

AD-A134 389

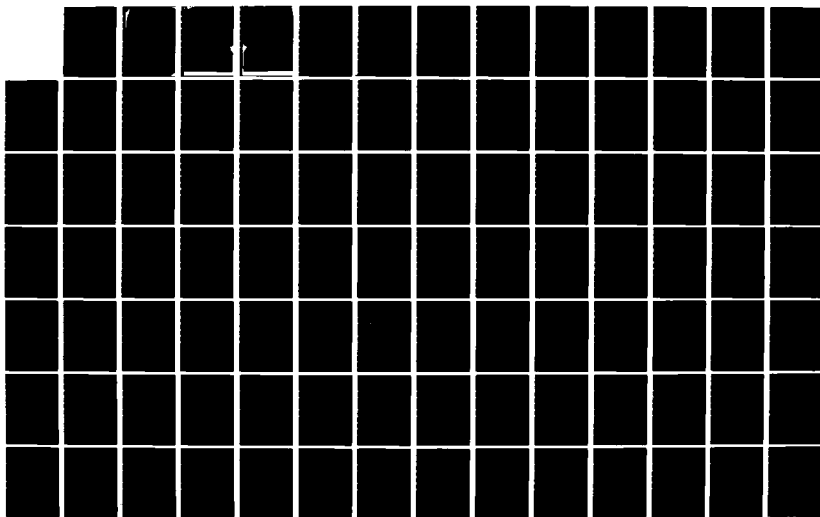
DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF
TOXICITY(U) OHIO STATE UNIV RESEARCH FOUNDATION
COLUMBUS P I FEDER ET AL. MAY 81 DAND17-79-C-9150

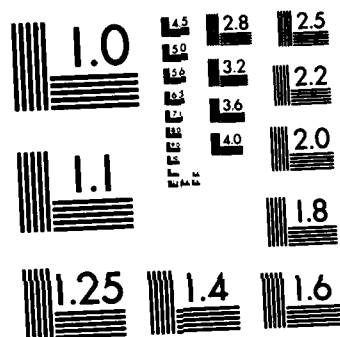
1/5

UNCLASSIFIED

F/G 6/28

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A134309

AD

DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY

Final Report

By

Paul I. Feder
Department of Statistics, The Ohio State University and
Applied Statistics Group, Battelle Columbus Laboratories

William J. Collins
Department of Entomology, The Ohio State University

(15 September 1979 - 31 October 1980)

Supported by

U.S. Army Medical Research and Development Command
Fort Detrick, Frederick, Maryland 21701

Contract No. DAMD 17-79-C-9150

The Ohio State University Research Foundation
1314 Kinnear Road
Columbus, Ohio 43212

DTIC
ELECTE

NOV 2 1983

A

Contract Officer's Technical Representative: Paul H. Gibbs
U.S. Army Medical Bioengineering Research and Development Laboratory
Fort Detrick, Frederick, Maryland 21701

Approved for public release; distribution unlimited

The findings in this report are not to be construed as an official
Department of the Army position unless so designated by other
authorized documents.

83 : 09 07 005

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A134	309
4. TITLE (and Subtitle) DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY		5. TYPE OF REPORT & PERIOD COVERED Final Report 9/15/79 - 10/31/80
		6. PERFORMING ORG. REPORT NUMBER 761860/712404
7. AUTHOR(s) Paul I. Feder William J. Collins		8. CONTRACT OR GRANT NUMBER(s) DAMD 17-79-C-9150
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 621102A.3E161102BS04. 00.048
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Medical Research and Development Command Medical Bioengineering Research and Development Lab Fort Detrick Frederick, Maryland 21701		12. REPORT DATE May, 1981
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 415
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) chronic aquatic toxicity tests multiple comparisons standardized data reporting dose response estimation tests of hypothesis computer programs confidence intervals statistical analysis fathead minnows		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The present investigation considers aspects of the planning of, reporting results from, and the statistical analysis of data arising from chronic aquatic tests of toxicity with fathead minnows. The discussion emphasizes analyses and data displays for the interpretation of qualitative mortality and abnormality data and quantitative weight data. The report consists of nineteen sections. Section I-IV describe the literature search, argue for and present a set of standardized data reporting		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (Abstract) - continued

→ formats, and contain some suggestions for further improving and standardizing the test procedure. Section V briefly describes the early life stage testing procedure and the resulting data. Sections VI-XIX discuss various topics pertaining to the statistical analysis of such data and to statistical aspects in the design of toxicity tests. Techniques discussed include preliminary graphical displays, tests for homogeneity among tanks within treatment groups, adjustments to account for tank to tank heterogeneity within groups, outlier detection procedures, overall tests for treatment effects, multiple comparisons and pairwise confidence intervals of treatment groups with the control group, analysis of variance, regression analysis and parametric and nonparametric dose response curve fits along with associated estimates of "safe concentrations". All suggested procedures are illustrated with real examples based on aquatic toxicity data. Several newly developed, specialized computer programs are described to carry out nonstandard analyses. ↗

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF TOXICITY

Final Report

By

Paul I. Feder

Department of Statistics, The Ohio State University and
Applied Statistics Group, Battelle Columbus Laboratories

William J. Collins

Department of Entomology, The Ohio State University

(15 September 1979 - 31 October 1980)

Supported by

U.S. Army Medical Research and Development Command
Fort Detrick, Frederick, Maryland 21701

Contract No. DAMD 17-79-C-9150

The Ohio State University Research Foundation
1314 Kinnear Road
Columbus, Ohio 43212

Contract Officer's Technical Representative: Paul H. Gibbs
U.S. Army Medical Bioengineering Research and Development Laboratory
Fort Detrick, Frederick, Maryland 21701

Approved for public release; distribution unlimited

The findings in this report are not to be construed as an official
Department of the Army position unless so designated by other
authorized documents.

EXECUTIVE SUMMARY

Purpose

During the past several decades problems of environmental contamination have become increasingly important both from the scientific and the legal standpoints. In recent years a great deal of attention has been directed to the potential toxicity to aquatic organisms of chemicals discharged into water bodies.

The U.S. Army, through activities such as munitions manufacturing, operates a number of plants that produce, consume, or discharge a variety of chemical substances. Some of these discharges enter bodies of water inhabited by various aquatic species. Thus the Army must provide the USEPA with safety data concerning the levels of such discharges and the possible extent of resulting surface water and ecological contamination. In order to develop such data the Army conducts both intramural and extramural programs of aquatic toxicity testing.

Considerable amounts of time, money, and manpower are expended by the Army in such aquatic toxicity testing programs. To make these programs more efficient and more effective, the need has been felt for a reexamination of some of the standard methods used. This has been especially true of statistical methods involved in the design of testing programs and the analysis of resulting data. This study is an effort to make some progress in those directions.

The results in this study indicate areas where the conduct of and the summarization and reporting of results from chronic aquatic toxicity tests can be further standardized and made easier to understand. A number of the statistical approaches and procedures discussed and/or developed in this study have not to the authors' knowledge been previously applied to aquatic toxicity data. These improved methods provide increased information, as compared with standard methodology, about the structure, relations, and anomalies in the data. They enhance the sensitivity of statistical analyses, so that greater precision of results can be obtained without increasing the amount of testing. In brief this study provides methods that should improve the reporting and statistical analysis of data from chronic aquatic toxicity tests. This will enhance the sensitivity of conclusions that can be derived from these tests, thereby increasing their efficiency.

Approach

All the statistical procedures discussed in this report are illustrated with examples based on real data from chronic tests with fathead minnows. At the outset of the study the principal investigators visited the USEPA Environmental Research Laboratory at Duluth to become oriented to the apparatus and procedures used in chronic toxicity tests. Discussions were held with various investigators concerning the details of their studies. Some of these investigators provided us with illustrative data to be used in our subsequent work.

A number of sets of experimental data were received from Duluth. The literature pertaining to chronic toxicity tests in general and to those tests in particular was reviewed and discussed between statistician and toxicologist. Based on an understanding of experimental procedures it was possible to then start considering the statistical aspects of the problems. The statistical procedures discussed in the body of this report represent a combination of methods taken from the statistical literature where appropriate or developed especially for aquatic toxicity testing applications where standard procedures were felt not to be the most appropriate.

Results

Arguments for the use of standardized fish stocks and standardized data reporting formats are presented. Aspects of the statistical analysis of toxicity data are discussed and the suggested procedures are illustrated with examples based on fish toxicity studies. Data analysis topics discussed include: graphical displays, preliminary tests of tank to tank heterogeneity within treatment groups, preliminary outlier detection tests, overall tests of heterogeneity in response rates across treatment groups, treatment group-control group pairwise multiple comparison procedures, the fitting of standard and nonstandard dose response curve models, analysis of variance and multiple regression analyses on quantitative responses, statistical power and estimation precision to be expected for various levels of sample size and suggestions for unequal allocation of experimental effort across treatment groups with greater effort expended on the control group and lower treatment groups.

Conclusions and Recommendations

1. The USEPA should revise and update the standard procedure for life cycle tests on fathead minnows.
2. Standardized data reporting sheets are a very useful adjunct to the categorization and analysis of chronic toxicity test data.

3. Guidelines on disposal of potentially hazardous effluent from chronic toxicity tests should be incorporated in the procedure.
4. Detailed procedures for chemical analysis and quality assurance of chemical data should be incorporated in the procedure.
5. Some of the "standard" methods currently used for analyzing data from aquatic toxicity tests can and should be modified. The data should first be graphed, outlying observations or groups of observations should be located and the reason for their aberrant behavior determined, and tests for heterogeneity among tanks within groups should be carried out. Based on the results of these preliminary inferences, the data should be modified or adjusted to account for possible heterogeneity or aberrant values before going on to the inferences of primary interest.
6. If hypothesis tests are to be used to compare the treatment group and control group responses they should be one sided tests which are sensitive to monotone alternatives, rather than overall analysis of variance type "shotgun" tests.
7. Multiple comparison procedures and confidence intervals procedures should be used to determine specifically which treatment groups have responses which differ from the control group responses and whether the differences are of a specified biological significance. Significance tests, by themselves, are not adequate to define an "MATC" (i.e., maximum acceptable tolerable concentration). Perhaps a confidence bound should be routinely constructed at the MATC to determine just how much worse than the control group the response at that concentration could conceivably be. In general, confidence intervals impart much more information than hypothesis tests and should be routinely used.
8. A good way to place monotone response structure on the problem, to smooth the data, and to convert a hypothesis testing problem into an estimation problem is to fit dose response curve models to the data and to define the "safe" concentration as that which results in no more than a specified increment in response from the control group. A number of nonstandard variants on the "standard" dose response models discussed in the literature may be useful. A nonparametric approach to dose response estimation is feasible, has been implemented in a computer program, and may be preferable on occasion to some of the standard parametric dose response models.

9. Statistical power and estimation precision depend both on the number of tanks run per group and the number of fish per tank. In the presence of substantial tank to tank heterogeneity the effective sample size may be more nearly the number of tanks than the number of fish. Thus in the presence of tank to tank heterogeneity, diminishing returns result from increasing the number of fish used without also increasing the number of fish tanks per group.
10. Under certain circumstances it is sensible to allocate experimental resources so that the control group and lower concentration groups receive more tanks and fish than the higher concentration groups. This results in greater inference sensitivity in the region of the MATC. Proportional diluters should be modified to permit such asymmetrical allocations of tanks, at the discretion of the investigator.
11. Statistical power or statistical precision goals should be stated as part of the protocol for each individual toxicity test and sample sizes should be determined accordingly.

PUBLICATIONS AND PRESENTATIONS SUPPORTED BY THE CONTRACT

Reports and Papers

1. Feder, P. I. and Sherrill, M.C., "A Computer Program to Calculate Nonparametric Lower Confidence Bounds on Safe Concentration in Quantal Response Toxicity Tests, "Technical Report No.215, Department of Statistics, The Ohio State University, Columbus, Ohio, September 1980.
2. Feder, P.I. and Willavize, Susan A., "EXAX2 -- A Computer Program to Compare Binomial Proportions, Program Description and Card Input Information, "Technical Report No.216, Department of Statistics, The Ohio State University, Columbus, Ohio, September 1980.
3. Feder, P.I. and Collins, W.J., "Considerations in the Design and Analysis of Chronic Aquatic Tests of Toxicity," Aquatic Toxicology and Hazard Assessment: Fifth Conference, ASTM STP 766, J.G. Pearson, R.B. Foster and W.E. Bishop, Eds., American Society for Testing and Materials, 1982, pp.32-68.

Presentations

1. "Considerations in the Design and Analysis of Chronic Aquatic Tests of Toxicity," Feder, P.I. and Collins, W.J. Fifth Annual Symposium on Aquatic Toxicology, American Society for Testing and Materials, Committee E35, October 1980.
2. "Design and Analysis of Aquatic Tests of Toxicity," Feder, P.I. 36th Annual Conference on Applied Statistics, American Society for Quality Control and American Statistical Association, December 1980.

PERSONNEL RECEIVING SUPPORT UNDER THE CONTRACT

Paul I. Feder, Department of Statistics

William J. Collins, Department of Entomology

Susan Willavize, Department of Statistics

Mary Sherrill, Department of Statistics

Gerald Skinner, Department of Statistics

Foreword

The investigators gratefully acknowledge the assistance and expert contributions to the project of the following graduate assistants: Susan Willavize and Mary Sherrill, Department of Statistics and Gerald Skinner, Department of Zoology. Charles Stephan of the USEPA Environmental Research Laboratory, Duluth, Minnesota was kind enough to arrange for us to talk to a number of investigators at his laboratory about the conduct of toxicity tests. Duane Benoit, Dave Defoe, Gary Holcombe, Fred Jarvinen, Gary Phipps and Robert Spehar of that laboratory and Gary Kimball, formerly at the University of Minnesota, sent us a number of data sets to be used to illustrate the statistical methodology discussed in the report. EPA, Environmental Research Laboratory - Duluth gave permission to use these data for illustrative purposes in the report. Paul Gibbs, Jerry Highfill, J. Gareth Pearson and William van der Schalie of the U.S. Army Medical Bioengineering Research and Development Laboratory provided guidance, suggestions, and coordination throughout the conduct of the study. Wayne Nelson kindly granted permission to reproduce his charts for confidence limits for the ratio of two Poisson means.

TABLE OF CONTENTS

	<u>Page</u>
Executive Summary	ii
Publications and Presentations	vi
Foreword	viii
Table of Contents	ix
Figures	xiv
Tables	xviii
Introduction	xix
 <u>Section</u>	
I. Assemble and Evaluate Information on Test Methods	1
II. Categorizing Data Sets	2
III. Organization of Data: Standardized Data Records	3
A. Composite of Experimental Conditions	3
B. Data Sheets for Separate Categories of Data	5
C. Transfer of Experimental Data to Standardized Data Sheets	9
IV. Analysis of and Comments on the Test Procedure	10
A. Physical System	10
B. Chemical System	10

<u>Section</u>	<u>Page</u>
V. Early Life Stage Toxicity Tests	11
A. Background	11
B. Data	11
VI. Preliminary Graphical Data Displays	18
VII. Testing for Tank to Tank Heterogeneity Within Treatment Groups	27
A. Background	27
B. Remarks on Some "Standard" Procedures	27
C. Separate Heterogeneity Tests Within Treatment Groups	32
VIII. EXAX2 -- A Computer Program to Test for Heterogeneity of Responses Among Tanks Within Treatment Groups	44
IX. Adjustments to Account for Tank to Tank Heterogeneity Within Treatment Groups	51
A. Background, Derivations, and Discussion	51
B. Illustrations	55
X. Outlier Detection Procedures	63
A. Background	63
B. Applications of Outlier Detection Procedures to Fish Toxicity Data	64
XI. Testing for Concentration Related Effects	88
A. Background	88
B. Chi Square Test of Homogeneity Across Treatment Groups	89
C. One Sided Tests of Homogeneity Across Treatment Groups	92

<u>Section</u>	<u>Page</u>
XII. Treatment Group vs Control Group Pairwise Multiple Comparison Procedures	116
XIII. Confidence Interval Procedures for Comparison of Treatment Group and Control Group Response Rates	120
A. Introduction	120
B. Method 1 Asymptotic Approach	122
C. Method 2 Exact, Small Sample Confidence Intervals	125
D. Method 3 Poisson Approximation	132
XIV. Dose Response Curve Estimation -- Probit Analysis	145
A. Introduction. Dose Response Estimation vs Hypothesis Testing	145
B. Probit Analysis Using SAS PROC PROBIT	147
XV. Dose Response Curve Estimation -- Maximum Likelihood Estimation by Nonlinear Least Squares Regression	177
XVI. Nonparametric Lower Confidence Bounds on Safe Concentrations	199
XVII. Analysis of Quantitative Responses	244
A. Preliminary Scatterplots	246
B. Outlier Detection Procedures and Testing for Tank to Tank Heterogeneity Within Groups	246
C. Multiple Comparison Procedures and Regression Analysis	251
XVIII. Experimental Design Considerations	267
A. Assumptions, Additional Variables to Measure, Numbers and Allocations of Tanks	267

<u>Section</u>	<u>Page</u>
B. Sample Size and Power Considerations for Qualitative Survival Data	270
C. Expected Precision for Comparisons of Treatment Group and Control Group Survival Probabilities	277
D. Power Calculations for Quantitative Weight Response	279
E. Unequal Allocations of Testing Effort Among Treatment Groups	284
XIX. Design and Analysis Considerations for Full Life Cycle Tests Distinct from Those for Early Life Stage Tests	288
References	290
Appendix AII* Early Life Stage Data Sets Used as Examples in the Body of the Report	294
Appendix AVII* Theoretical Bases of Chi Square and Probit Based Tests of Tank to Tank Variation Within Treatment Groups	306
Appendix AVIII.1* Output from EXAX2 Computer Program	308
Appendix AVIII.2* Description and Instructions for Use of EXAX2 Computer Program	345
Appendix AX* Theoretical Bases of Suggested Outlier Detection Transformations	362
Appendix AXV* Confidence Interval on Concentration That Corresponds to a Given Level of Increase in Resonse Over Control Group Response	365

* Appendices are numbered so as to indicate their relation to corresponding section numbers in the body of the report.

<u>Section</u>	<u>Page</u>
Appendix AXVI.1* Confidence Bounds on Binomial Probabilities	368
Appendix AXVI.2* Description and Instructions for use of Nonparametric Dose Response Estimation Computer Program	369

* Appendices are numbered so as to indicate their relation to corresponding section numbers in the body of the report.

FIGURES

<u>Section</u>		<u>Page</u>
V	1. Standardized data reporting format for results of early life stage test -- compound C -- DeFoe	13
	2. Computerized data format for results of early life stage test -- compound C -- DeFoe	16
VI	1. DeFoe compound C percentage embryo mortality by treatment group	19
	2. DeFoe compound C percentage fry mortality by treatment group	20
	3. DeFoe compound C percentage embryo abnormality by treatment group	21
	4. DeFoe compound C percentage fry abnormality by treatment group	22
	5. Holcombe and Phipps compound D embryo mortality by treatment group	23
	6. Holcombe and Phipps compound D fry mortality by treatment group	24
	7. Holcombe and Phipps compound D embryo abnormality by treatment group	25
	8. Holcombe and Phipps compound D fry abnormality by treatment group	26
VII	1. Probit fit to individual responses -- DeFoe compound C (excluding control group)	34
	2. Probit fit to responses pooled across tanks within groups -- DeFoe compound C (excluding control group)	35
	3. Chi square test of homogeneity based on individual responses -- DeFoe compound C fry mortality	36
	4. Chi square test of homogeneity based on responses pooled across tanks within groups -- DeFoe compound C fry mortality	37

<u>Section</u>		<u>Page</u>
	5-10. Chi square test of homogeneity of percent fry mortality in groups 1-6. Data from Holcombe and Phipps compound D	38
X	1. Normal probability plot of standardized values based on DeFoe compound C embryo mortality data -- to detect outlying responses	83
	2. Normal probability plot of standardized values based on DeFoe compound C fry mortality data -- to detect outlying responses	84
	3. Normal probability plot of standardized values based on Holcombe and Phipps compound D embryo mortality data -- to detect outlying responses.	85
	4. Normal probability plot of standardized values based on Holcombe and Phipps compound D fry mortality data -- to detect outlying responses	86
	5. Normal probability plot of standardized values based on Jarvinen compound B embryo mortality data -- to detect outlying responses	87
XI	1. SAS PROC FREQ output from chi square test for homogeneity across groups applied to fry mortality data from Holcombe and Phipps test on compound D	101
	2-7. EXAX2 output -- chi square test of homogeneity across groups	102
	8. Input information required for program BMDP1F	108
	9-11. BMDP1F output from one sided measure of association test on artificial data	109
	12-15. BMDP1F output on real data	112
XIII	1-2. EXAX2 output from calculation of exact, small sample confidence intervals on the odds ratios between treatments and control	137

<u>Section</u>		<u>Page</u>
	3-5. Poisson two sample confidence interval charts -- 90, 95, 99 percent (from Nelson, 1969)	142
XIV	1. Schematic representation of dose response curve formulation	159
	2. Basic data	160
	3-6. Output from PROC PROBIT fit to DeFoe fry mortality data -- arithmetic concentration	161
	7-8. PROC PROBIT fit to DeFoe fry mortality data -- no background correction	165
	9. PROC PROBIT fit to DeFoe fry mortality data -- logarithmic transformation of concentration	167
	10. Basic data	168
	11-14. Output from PROC PROBIT fit to Holcombe and Phipps compound D fry mortality data -- arithmetic concentration, logarithmic concentration	169
	15-18. Plots of observed and predicted based on probit fit -- arithmetic concentration, logarithmic concentration	173
XV	1. Holcombe and Phipps compound D mortality and concentration data -- pooled	190
	2-4. Output from SAS PROC NLIN applied to Holcombe and Phipps data -- untransformed concentration	191
	5-7. Logit fit to Holcombe and Phipps data using SAS PROC NLIN -- untransformed concentration	194
	8. Use of SAS PROC NLIN to fit nonstandard dose response model to data	197
	9. Output from SAS PROC NLIN to fit nonstandard dose response model to data	198
XVI	1. Schematic description of nonparametric dose response estimation procedure -- Holcombe and Phipps compound D fry mortality data	241
	2-3. Benoit compound A -- fry mortality vs arithmetic concentration, logarithmic concentration	242

<u>Section</u>		<u>Page</u>
XVII	1. Average weight per tank vs group	254
	2. Standard deviation of weight per tank vs group	255
	3. Average weight per tank vs concentration	256
	4. Average weight per tank vs log concentration	257
	5-6. Analysis of variance fit of individual weights (logarithmic weights) on group and tank within group. Holcombe and Phipps compound D fry mortality data	258
	7. Expected mean squares in analysis of variance fits to Holcombe and Phipps fry mortality data	260
	8. Analysis of variance fit of average weights per tank on group. Holcombe and Phipps compound D fry mortality data	261
	9. Residuals vs group	262
	10. Logarithmic residuals vs group	263
	11. Normal probability plot of residuals from untransformed weights	264
	12. Normal probability plot of residuals from logarithmic weights	265
	13. Regression fit of weight gain on $\log(1 + \text{CONC})$. Holcombe and Phipps compound D fry mortality data	266
XVIII	1. Effective sample size per group vs sample size for various values of θ	287

TABLES

<u>Section</u>		<u>Page</u>
VII	1-2 DeFoe compound C fry mortality data -- probit fit to groups 2, 3, 4, 5, 6, -- individual tanks, pooled tanks within groups	30
IX	1. Effective sample sizes and responses in Holcombe and Phipps compound D fry mortality data after adjustment for tank to tank heterogeneity	58
	2. Effective sample sizes and responses in Jarvinen compound B embryo mortality data after adjustment for tank to tank heterogeneity	62
XVIII	1. Power of one sided pairwise comparisons of survival rates between control group and treatment groups. $\alpha = 0.05$	272
	2. Expected half lengths of 95 percent two sided confidence intervals for comparisons between treatment group and control group response rates	277
	3. Power of one sided pairwise comparisons of average weight gains between control group and treatment group. $\alpha = 0.05$	281

INTRODUCTION

During the 1960's and 1970's, environmental contamination in general and water pollution specifically, became increasingly important as legal and scientific problems. Regulatory agencies needed scientific data to support the notion that a problem existed and also needed factual information for establishing tolerance limits for levels of chemical discharges into surface waters. From that need evolved numerous standard toxicity tests, including a test to determine the long term effects of toxicants on a representative fish, the chronic toxicity test with fathead minnows.

Aquatic toxicologists and biologists evolved and refined an effective fish toxicity test. As data were analyzed and experiments designed, the statistical considerations evolved to a more complex level. It became clear that some of the statistical procedures needed for the design of toxicity tests and for the analysis of chronic toxicity data may be novel or unique and should be developed specifically for fish toxicity tests.

Operational activities of the U.S. Army (e.g. munitions manufacture) involve the production, use, and/or discharge of a variety of commercial chemicals. Safety data must be provided to USEPA concerning surface water contamination due to discharges of chemical intermediates or the final product. In-house research of the U.S. Army with the standard fish chronic toxicity test highlighted the need for a reexamination of the standard procedures, especially regarding statistical techniques. The main goals of this project are to suggest statistical procedures for analyzing data arising from such toxicity tests, to provide recommendations for a more accurate, reliable standard procedure, and to facilitate research in aquatic toxicology in general.

This project was initiated as an interdisciplinary investigation of the EPA chronic toxicity test for fathead minnows. It combined the efforts of a toxicologist, a fish specialist, and statisticians. The biologists functioned as advisors to the statisticians in regard to the characteristics and limitations of the test animal, test procedures, and chronic toxicity data and evaluated the test procedure from a toxicological viewpoint.

The statisticians developed procedures for data storage, transformation and analysis, scrutinized published statistical techniques for their applicability to fish toxicity data, and devised new statistical methods for analyzing data from fish toxicity tests when they were felt to be more applicable than the standard methods.

This final report is the synthesis of a one year effort. It discusses both biological and statistical aspects of the planning, conduct, reporting, and data analyses associated with toxicity tests on fathead minnows.

Arguments for the use of standardized fish stocks and standardized data reporting formats are presented. Aspects of the statistical analysis of toxicity data are illustrated with examples based on fish toxicity studies. Data analysis topics discussed include: graphical displays, preliminary tests of tank to tank heterogeneity within treatment groups, preliminary outlier detection tests, adjustments in analysis procedures due to tank to tank heterogeneity, overall tests of heterogeneity in response rates across treatment groups, treatment group-control group pairwise multiple comparison procedures, the fitting of standard and nonstandard dose response curve models, analysis of variance and multiple regression analyses on quantitative responses, statistical power and estimation precision to be expected for levels of sample size and suggestions for unequal allocation of experimental effort across treatment groups with greater effort expended on the control group and lower treatment groups.

It is hoped that the results obtained in this study will contribute to better, more reliable toxicity tests and data analyses. This in turn should provide improved tools for the regulation of toxic chemicals in aquatic environments and should suggest fertile areas for further study and development.

I. ASSEMBLE AND EVALUATE INFORMATION ON TEST METHODS

Although the scope of work specified that design and analysis of chronic toxicity tests would be researched, it became clear early in the project that many aspects of statistical analysis could be pursued using exemplary data from early life stage tests. Consequently, the review of test methods and our literature search included both chronic life-cycle tests and early life stage tests.

Three literature sets were searched by computer using appropriate key words (fathead minnow, toxicity, chronic tests, etc.): Mechanized Information Center, The Ohio State University; Oak Ridge National Laboratory, Oak Ridge, Tenn.; Ohio Environmental Protection Agency, Columbus, Ohio. The latter search encompassed 13 major data base searches (Amer. Chem. Soc., Biol. Abstracts, etc.). The hundreds of citations received were reviewed for relevancy and the important ones were abstracted and filed. Reprints of copies or articles, 64 in all, that were considered to be directly related to future tasks were assembled, catalogued and reviewed in detail.

These publications provided information on the fathead minnow in regard to biology, life cycle events, duration of developmental stages, nutritional information and reproductive characteristics. The papers on test methods provided details of variation in design among investigators and a large amount of experimental toxicity data for reference and further discussion.

Papers and technical reports from E.P.A.-Duluth describing the apparatus [1] and procedure [2] for chronic toxicity tests were reviewed and studied in detail in order to understand the method of exposure, physical arrangement of the delivery system and important variations among investigators, e.g. the syringe delivery method of DeFoe [3]. The literature search and research paper perusal was essential for the toxicology group to authoritatively interpret biological factors, experimental data or test methods in discussions with the statistical group or to suggest limitations in design due to the animal or technique.

II. CATEGORIZING DATA SETS

Early in the project, a number of sets of experimental data from early life stage toxicity tests and chronic life cycle toxicity tests were received from and discussed with researchers in EPA-Duluth. Various aspects of these data were reviewed with the statistical group to clarify experimental procedures such as types of measurements and how they were acquired, if measurements were destructive or nondestructive, the replication of experiments and variation of chemical concentration in the delivery system. These discussions brought out the need for standardized data reporting sheets so that data could be accurately categorized and recorded in a systematic manner for entry into the computer and statistical analysis. A separate section on standardized data sheets is included in this report.

The examples in this report are based on some of these experimental data sets. In particular the data sets from early life stage toxicity tests by

Benoit - compound A
DeFoe - compound C
Holcombe and Phipps - compound D
Jarvinen - compound B

are used. These data sets are listed in Appendix AII.

III. ORGANIZATION OF DATA: STANDARDIZED DATA RECORDS

The early life stage data and chronic life cycle toxicity data sets were supplied by six different investigators. Each set of chronic data was received in a unique format. Each set of data was reviewed for the experimental procedure (if available) in order to accurately categorize the data for storage on computer and subsequent statistical analysis. Routine questions, e.g. how many days of exposure, and more complex questions, e.g. are replicate tests genuine replications, were not easily resolved by a review of the data sheets, nor was the comparability of the same categories of data in similar experiments among the array of investigators. Standardized data records have merit if they are sufficiently versatile to meet most needs, clearly summarize the exposure conditions and facilitate transformation, computer storage, and statistical analysis of raw data. The latter task is often done by an individual who is not an expert in biological research and unfamiliar with operational details of toxicity tests.

Good laboratory practice regulations (GLP) have been adopted by FDA for nonclinical laboratory studies [4] to assure the quality of data in support of product safety decisions. One component of the GLP deals with specific record-keeping practices for experimental data. The advantages of these required record-keeping practices have been discussed in regard to vertebrate experiments [5] and would apply equally as well to fish toxicity data to the benefit of investigators, statisticians, and regulatory agencies.

Although there is some variation in the design of fish toxicity tests, certain features are almost universal. For example, in a chronic toxicity test, a flow-through apparatus is always used and standard measurements include hatchability of embryos, fish length and weight, survival (mortality), and spawning data. Consequently, standardized recording sheets could be devised for summarizing exposure methods and experimental data.

Our standardized reporting sheets have two components: (a) a descriptive section summarizing the conditions of the experiment with code words or letters to categorize or define data for the statistician, and (b) the raw data record sheet with no calculations or transformations. These record sheets have been designed for data of early life stage tests or chronic life cycle tests.

A. Composite of Experimental Conditions

1. Investigator _____
2. Toxicant _____; Source _____ and _____ % purity
3. Starting Date of Test (Day Zero): ____ / ____ / ____.
- D M Y

4. Selection of embryos: _____ Embryos selected randomly.

_____ Embryos examined, only viable
eggs incubated.

_____ Other (specify).

5. Are embryos from paired matings? Yes or no.

6. Fish I.D.: for use only with paired matings; use a unique identifier here.

7. Generation of embryos or fish? Zero or first?

(Note: In chronic tests, some investigators refer to spawnings of exposed adults as first generation embryos, others, second generation. We define zero generation as any stage or form used to start a test and any stages during that generation, including adults. First generation is any of the stages following zero generation adults. One may argue against this system on a biological basis, but it distinguishes between the same stage of separate generations. With this system there is no second generation in a standard chronic life cycle test.)

8. Nominal Concentration: identify tanks by nominal concentration of toxicant (mg/l, ug/l), "Solvent control" by "S", "water only control" by "W".

9. Identify replicate tanks in Nominal Concentration column by "REP"
e.g. 0.25 REP; identify equivalent tanks by "EQ", e.g. 0.25 EQ.

(Note: It should always be clear in original data sheets the relation of replicate tanks of the same nominal concentration, a crucial factor in deciding what statistical procedures to use. We define replicate as the simultaneous exposure of fish to similar concentrations of a chemical that are delivered independently, i.e. two tanks containing nominal concentrations that originate from separate syringes in the delivery system are replicates. We define equivalent as the simultaneous exposure of fish to the same concentration of a chemical that is delivered from a common origin, i.e. groups of fish in several screened compartments of the same tank are equivalent groups, as are fish in different tanks supplied equally by tubing that is split after the final dilution. Replicate ("repeated experiment") does not distort the conventional meaning of that term. The choice of equivalent ("equal in quantity") was the best approximation of what occurs.

10. Tank I.D.: identify multiple tanks of the same nominal concentration and type (REP or EQ) by capital letters, e.g. Rep-A, Rep-B, etc.

No. days since day zero	Nominal Concentration	
	Tank I.D.	
	Initial No.	
_____ days	Alive Alive, normal Dead Lost Total for Interval	
_____ days	Alive Alive, normal Dead Lost Total for Interval	
_____ days	Alive Alive, normal Dead Lost Total for Interval	

2. Fry length or weight

- a. Experimental conditions: Use entries, 1, 2, 3, 7, 8, 9, 10 and 14 from Part A to summarize the conditions of the experiment.
- b. Data Sheet

Investigator _____

Toxicant _____ % purity _____

Measurements taken days after first exposure of this stage.

Starting Date of Test (day zero): / /

 D M Y

Nominal Conc.	Tank I.D.	Length (mm) or Weight (mg) of Individual Fry													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14

3. Hatchability of embryos

- a. Experimental conditions: Use entries 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 from Part A to summarize the conditions of the experiment.

- b. Data Sheet

Investigator _____

Toxicant _____ % purity _____

Starting date of Test (day zero): / /
D M Y

- b. Data Sheet**

Investigator _____

Toxicant _____ % purity _____

Starting date of test (day zero): / /

 D M Y

Nominal conc.	Tank I.D.	Fish I.D.	No. days since day zero	Length (mm)	Weight (mg)	Sex (m, f, im)
------------------	--------------	--------------	----------------------------	----------------	----------------	-------------------

7. A separate sheet that need not be standardized should be attached to the data records summarizing important conditions, i.e. ph, temperature, photo period, flow rates, type of food and feeding schedule, etc. and any limits of conditions that vary during the test.

C. Transfer of Experimental Data to Standardized Data Sheets.

After considerable debugging of the data sheets and several trials with actual data, the experimental data were transferred to data sheets for storage on the computer.

IV. ANALYSIS OF AND COMMENTS ON THE TEST PROCEDURE

The published procedure for chronic toxicity tests [2] has not been revised since 1972. Since that time considerable research on the test per se has been done at EPA - Duluth and elsewhere to improve reliability, reproducibility and accuracy. For example, some conditions specified in the 1972 procedure may be replaced by improved techniques, e.g. handling and selection of embryos, use of paired spawnings, etc. Those changes that could be incorporated as improvements in the test procedure should be made and a revised version published.

Following are comments about specific sections of the procedure, using the number and letter designations of the procedure [2] as a reference.

A. Physical System

4. Flow Rate

Recent USEPA regulations [7] designate certain chemicals as hazardous wastes if and when they are discarded. Guidelines and recommendations on the treatment (clean up) of experimental tank effluent should be included for a test system containing potentially toxic chemicals in ten or more tanks, changing 6 to 10 tank volumes/24 hours in each tank, all operating continuously for months.

14. Where surface water or municipal water is used, a filter system should be considered.

B. Chemical System

2. Measurement of toxicant concentration and

5. Methods

A much more detailed procedure should be incorporated in this section in conjunction with a carefully formulated standardized reporting sheet. The essence of this suggestion resides in the absolute need to know the limits of chemical concentration changes and to have assurance that the chemical analysis data are reliable.

V. EARLY LIFE STAGE TOXICITY TESTS

A. Background In recent years there has been some movement in the direction of developing toxicity tests that provide much of the information relating to chronic and sublethal toxicant effects that is obtainable from full life cycle tests, yet which require far fewer resources of time, space, cost and which are simpler to carry out and analyze. To accomplish these aims, the use of early life stage toxicity tests has become more common. For fathead minnows such early life stage tests require about thirty days of effort as compared with 250 to 300 days for full life cycle tests. This permits a great many more compounds to be tested.

A number of guidelines for conducting early life stage tests in a standardized manner have been proposed [8, 9, 10]. In these tests, organisms are exposed during part of the embryonic stage, throughout the larval stage, and during part of the juvenile stage. The rationale is that this represents the period of greatest sensitivity of the fish, and so chronic and sublethal toxicant effects will be revealed.

In one version of the test, groups of recently fertilized fish embryos are placed in embryo cups within test chambers. There are generally five or more toxicant groups and one or more control groups. Each (treatment or control) group consists of two or more replicate test chambers. The embryos are kept on test until they hatch (about 5 to 7 days), at which point the live, normal larvae are thinned to the desired number per tank and these are kept on test for about an additional four weeks, at which point the test is terminated. In a variant on this approach, the embryos are thinned after just two days on test. After hatch, all of the live larvae are released into the test chambers for the rest of the test. This avoids handling the newly hatched larvae at a time when they are most sensitive to the toxicant.

B. Data The data recorded in such early life stage tests include number of embryos per embryo cup, number of embryos hatched live and hatched normal, number of fry in each embryo cup after thinning, number of fry live at end of test and number normal, individual weights of all fry alive at end of test, and periodic toxicant concentration measurements within each tank.

Standardized data reporting sheets that facilitate the interpretation of test results and the communication of these results among investigators and laboratories have been developed by investigators at USEPA - Duluth. They have been kind enough to supply us with such sheets from about twenty early life stage tests (personal communication). Figure V. 1 illustrates such a basic data reporting sheet based on the test of compound C carried out by DeFoe.

Page 1 contains embryo and fry survival and normality data and a diagram showing the test layout. We see that in this test there was a single control group (1), five treatment groups (2 to 6), two test chambers per group, and a single embryo cup per test chamber. Page 2 contains individual weight measurements on all the fry that survived the test. Page 3 contains the results of the individual toxicant concentration measurements made in each chamber periodically throughout the test. We have found these data reporting sheets to be very easy to understand and very useful.

In order to work with the data it was necessary to put them into computer readable form. The approach that we took to accomplish this is illustrated in Figure V.2 for the data from the test of compound C by DeFoe. The three types of data -- survival, weight, and toxicant concentration-- are represented in three "card types." The data for each "card type" are listed in Figure V.2. Some applications call for use of just one card type while others call for use of two or more card types. The first six entries on each card are the same across card types -- treatment group (col 2), replicate designation (col 4), card type (col 6), card member (cols 7-8), investigator code (cols 9-10), test code (cols 11-12). This provides enough information to sort the cards by investigator, experiment, type, group, and sequence should the data become disarranged. Card type 1 (survival data) contains in addition number of embryos tested (cols 16-20), number hatched live (cols 21-25), number of fry tested (cols 31-35), number live at end of test (cols 36-40), number normal at end of test (cols 41-45). Card type 2 (weight data) contains number of weights recorded from that particular chamber (cols 14-15), individual weights (5 cols per weight, up to 13 weights per card). Card type 3 (toxicant concentration) contains month (cols 16-17), day (cols 18-19), year (cols 20-21), toxicant concentration (cols 32-38) -- one determination per card. At the head of each type of information several lines of descriptive text are given. This text is informative when the data are printed out but is skipped over for purposes of analysis.

We have found this data organization to be easy to prepare, easy to maintain, and easy to use. Such data files represent the "basic data" for all subsequent analyses discussed in this report.

INVESTIGATOR: DeFoe
TEST: Compound C

DRAW THE PHYSICAL LAYOUT SHOWING LOCATIONS OF ALL EMBRYO CUPS AND TEST CHAMBERS

Tox.	Doser

0 incoming H₂O
* Embryo Cup

Figure V.1 Standardized data reporting format for results of early life stage test --
compound C -- Defoe

2

INVESTIGATOR: DeFoe

TEST: Compound C

1 (CENTRAL)				2 (LOWEST)				3				4				5				6 (HIGHEST)			
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
158	011			124	091			171	170			097	115			037	029			0	0		
109	147			143	192			148	153			133	111			056	050			0	0		
106	172			012	195			095	151			088	070			072	038			↓	↓		
170	173			108	214			179	165			122	099			047	055						
166	174			141	198			086	131			187	138			056	035						
179	151			142	187			160	134			162	084			053	033						
124	200			101	095			128	149			089	132			037	057						
132	100			124	220			157	152			135	114			060	036						
068	124			162	116			179	162			165	064			044	066						
085	238			129	164			153	137			144	102			007	045						
118	173			153	162			102	120			122	090			028	050						
180	145			187	200			146	128			152	091			047	068						
092	170			128	190			137	131			073	131			049	003						
148	180			181	161			131	100			183	177			037	045						
163	165			149	206			158	096			144	063			004	038						
174	150			126	200			144	129			145	124			059							
138	089			103	190			152	152			145	090										
153	230			103	192			112	047			139	108										
172	135			192	137			177				226	119										
138	071			080	181			198				150											
												0%											

Figure V.1 (Cont)

3

INVESTIGATOR: DeFoe

TEST: Compound C

(Miles)

MEASURED CONCENTRATIONS OF TOXICANT

DATE	1 (CONTROL)			2 (LOWEST)			3			4			5			6 (HIGHEST)		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
9-5-77	.068			1.95			7.97			14.12			43.71			135.7		
9-6		.078			1.78			5.22			13.55			48.2			144.75	
9-7	.065			1.82			5.66			13.32			43.15			112.47		
9-10	.074			2.54			4.81			16.87			55.20			153.25		
9-11		.054			2.34			7.06			17.12			49.50			182.75	
9-12	.116			2.24			4.28			12.35			52.20			156.25		
9-13		.045			2.15			8.40			16.57			51.50			153.0	
9-14	.068			1.84			6.16			12.89			42.46			131.0		
9-17		.018			1.94			4.73			14.76			48.24			X	
9-19	.023			1.83			6.55			14.19								
9-20		.042			1.88			6.20			13.78			44.28			X	
9-27		.026			1.54			4.54			14.37			47.70			X	
9-28	.024			1.76			4.73			14.99			48.36			X		
10-1		.024			2.0			6.77			15.10			51.79			X	
10-2	.033			2.02			6.08			13.88								
10-3		.033			2.01			6.74			17.14			51.63			X	
														51.88			X	

Figure V. 1 (Cont)

INVESTIGATOR: DEFOE (01), TEST: COMPOUND C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 2 REPS EA (A,B)
 INDIVIDUAL WEIGHTS (MG) OF ALL FISH ALIVE AT END OF TEST (2 CARUS OR LESS/CELL)
 NUMBER OF WEIGHTS & LIST OF WEIGHTS

1	A	1	10201	20	158	159	176	166	179	124	132	608	085	118	180	292
1	B	2	10201	20	149	163	174	138	153	172	138					
1	B	3	10201	20	111	147	172	173	174	151	200	100	124	238	173	145
1	B	4	10201	20	130	165	150	089	230	135	071					
1	B	5	10201	20	144	143	012	108	141	142	101	124	162	129	153	197
2	A	2	10201	20	181	149	126	153	103	193	080					
2	A	3	10201	20	091	192	195	214	158	187	095	220	116	164	162	200
2	A	4	10201	20	161	206	200	190	192	197	131					
2	A	5	10201	20	171	148	095	179	086	160	128	157	179	153	102	146
3	A	2	10201	20	131	158	144	152	112	177	198					
3	A	3	10201	19	170	153	151	165	131	134	149	152	162	137	120	128
3	A	4	10201	19	100	096	129	156	047							
3	A	5	10201	21	097	133	068	122	169	162	089	135	145	144	122	152
4	A	2	10201	21	183	144	145	145	139	226	150	046				
4	A	3	10201	19	115	111	070	094	138	084	132	114	064	102	090	091
4	A	4	10201	19	177	063	124	090	108	119						
4	A	5	10201	16	037	056	072	047	056	053	037	260	244	007	028	047
5	A	2	10201	16	057	054	059									
5	A	3	10201	15	029	050	038	055	035	033	057	036	066	045	150	068
5	A	4	10201	15	045	038										
5	A	5	10201	00												
6	B	2	10201	00												

INVESTIGATOR: DEFOE (02), TEST: COMPOUND C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 2 REPS EA (A,B)
 # EMBRYOS TESTED, # LIVE AFTER HATCH, # NORMAL FRY AFTER HATCH
 # FRY TESTED, # ALIVE AT END, # NORMAL AT END

1	A	1	10201	50	28	28	20	20	20							
1	B	1	10201	50	31	31	20	20	20							
2	A	1	10201	50	18	17	20	20	20							
2	B	1	10201	50	31	31	20	20	20							
3	A	1	10201	50	10	10	20	20	20							
3	B	1	10201	50	33	33	20	19	19							
4	A	1	10201	50	34	34	21	21	21							
4	B	1	10201	50	29	26	21	14	14							
5	A	1	10201	50	28	28	20	16	16							
5	B	1	10201	50	33	33	20	15	15							
6	A	1	10201	50	31	00	20	00	00							
6	B	1	10201	50	31	00	20	00	00							

INVESTIGATOR: DEFOE (02), TEST: COMPOUND C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 2 REPS EA (A,B)
 MEASURED CONCENTRATIONS OF TOXICANT (MG/L NO UNITS GIVEN)

1	A	3	10201	090579	000.068											
1	A	3	20201	090779	000.065											
1	A	3	30201	091079	000.074											
1	A	3	40201	091279	000.116											
1	A	3	50201	091479	000.068											
1	A	3	60201	091679	000.123											
1	A	3	70201	091879	000.024											
1	A	3	80201	092079	000.033											
1	B	3	10201	090679	000.078											
1	B	3	20201	091179	000.054											
1	B	3	30201	091379	000.045											
1	B	3	40201	091779	000.018											
1	B	3	50201	092079	000.042											
1	B	3	60201	092779	000.020											
1	B	3	70201	100179	000.024											
1	B	3	80201	100379	000.033											
2	A	3	10201	090579	001.95											
2	A	3	20201	090779	1.82											

Figure V.2 Computerized data format for results of early life stage test -- compound C -- DeFoe

2	A	3	30201	091079	2.54
2	A	3	40201	091279	2.24
2	A	3	50201	091479	1.84
2	A	3	60201	091979	1.54
2	A	3	70201	092479	1.46
2	A	3	80201	092779	2.02
2	A	3	90201	093679	1.78
2	A	3	00201	094179	2.34
2	A	3	10201	091379	2.15
2	A	3	20201	091779	1.94
2	A	3	30201	092079	1.88
2	A	3	40201	092779	1.54
2	A	3	50201	093179	2.01
2	A	3	60201	093279	2.01
2	A	3	70201	093579	2.47
2	A	3	80201	093779	5.66
2	A	3	90201	091079	4.81
2	A	3	00201	091279	4.28
2	A	3	10201	091479	6.16
2	A	3	20201	091979	6.55
2	A	3	30201	092479	4.73
2	A	3	40201	092779	6.49
2	A	3	50201	093679	6.22
2	A	3	60201	094179	7.06
2	A	3	70201	091379	8.40
2	A	3	80201	091779	4.73
2	A	3	90201	092079	0.20
2	A	3	00201	092779	4.54
2	A	3	10201	093179	6.59
2	A	3	20201	093279	6.74
2	A	3	30201	093579	14.17
2	A	3	40201	093779	13.32
2	A	3	50201	091079	16.87
2	A	3	60201	091279	17.35
2	A	3	70201	091479	12.89
2	A	3	80201	091979	14.19
2	A	3	90201	092479	14.49
2	A	3	00201	092779	13.88
2	A	3	10201	093679	13.55
2	A	3	20201	094179	17.12
2	A	3	30201	091379	16.57
2	A	3	40201	091779	14.76
2	A	3	50201	092079	13.73
2	A	3	60201	092779	14.37
2	A	3	70201	093179	15.10
2	A	3	80201	093279	14.14
2	A	3	90201	093579	43.71
2	A	3	00201	093779	43.15
2	A	3	10201	091079	55.80
2	A	3	20201	091279	42.20
2	A	3	30201	091479	42.40
2	A	3	40201	091979	46.11
2	A	3	50201	092479	49.36
2	A	3	60201	092779	51.03
2	A	3	70201	093679	48.01
2	A	3	80201	094179	48.50
2	A	3	90201	091379	51.57
2	A	3	00201	091779	48.24
2	A	3	10201	092079	44.20
2	A	3	20201	092779	43.00
2	A	3	30201	093179	51.49
2	A	3	40201	093279	51.83
2	A	3	50201	093579	135.9
2	A	3	60201	093779	117.47
2	A	3	70201	091079	154.25
2	A	3	80201	091279	150.75
2	A	3	90201	091479	131.75
2	A	3	00201	091979	144.75
2	A	3	10201	092479	122.75
2	A	3	20201	092779	153.00

VI. PRELIMINARY GRAPHICAL DATA DISPLAYS

Graphing the data is generally considered to be a good first step in analyzing data. Graphs provide insights into the structure of the data and reveal the presence of possibly unanticipated relations or anomalies in the data.

Figures VI.1 and VI.8 illustrate the kinds of information that can be obtained from preliminary plots. They illustrate percentage embryo and fry mortality and abnormality observed in early life stage tests on fathead minnows conducted by DeFoe with compound C and by Holcombe and Phipps with compound D. The tests each consist of a control group (1) and five treatment groups (2 - 6) with toxicant levels in roughly geometric progression. The DeFoe test was run with four chambers per group. The plotting symbol "A" represents a single response, "B" represents two coincident responses, etc.

Figures VI.1 and VI.5 reveal no trends in embryo mortality with increasing toxicant level in either test. Tank to tank variation within treatment groups appears to be approximately constant across groups except for a single tank in Group 2 (Figure VI.1) which has about 50 percent greater embryo mortality than all the other tanks in the test. It appears to be an outlier, i.e. its response does not seem to conform to the pattern of the bulk of the data.

Figures VI.2 and VI.6 show increasing trends in fry mortality with toxicant concentration in each test. This pattern is to be expected since the larvae are most toxicant sensitive shortly after hatching. In each test tank to tank variation within groups is greatest in the middle and least at the ends, in conformance with binomial theory. No outlying tanks are evident with respect to fry mortality. Note that in both tests the highest treatment groups experience 100 percent fry mortality.

Figures VI.3 and VI.7 exhibit embryo abnormality in the two tests. They are strikingly similar. In the control groups and the four lowest concentration groups there is little or no abnormality among newly hatched live larvae. However in the highest concentration groups there is 100 percent abnormality among newly hatched live larvae. It thus appears that very high concentrations of each of these toxicants will penetrate the embryo.

Figures VI.4 and VI.8 exhibit fry abnormality in the two tests. In brief, there is none. After 32 days the fry have either died or are normal. Recall that the highest toxicant groups experience 100 percent mortality and so there is no abnormality data to plot.

PLOT OF PMRTEM_B*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

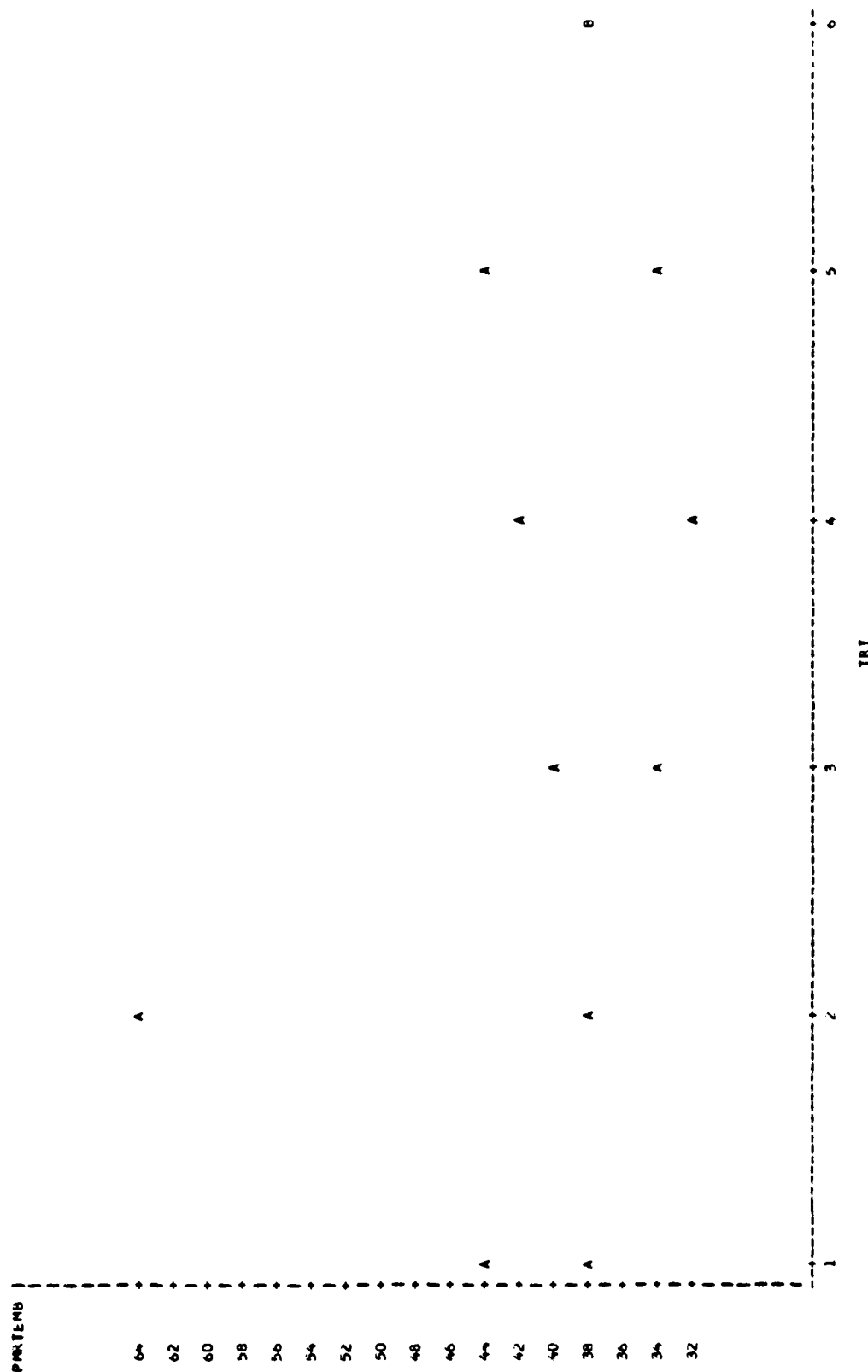


Figure VI.1 DeFoe compound C percentage embryo mortality by treatment group

PLOT OF PMKTRFY*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

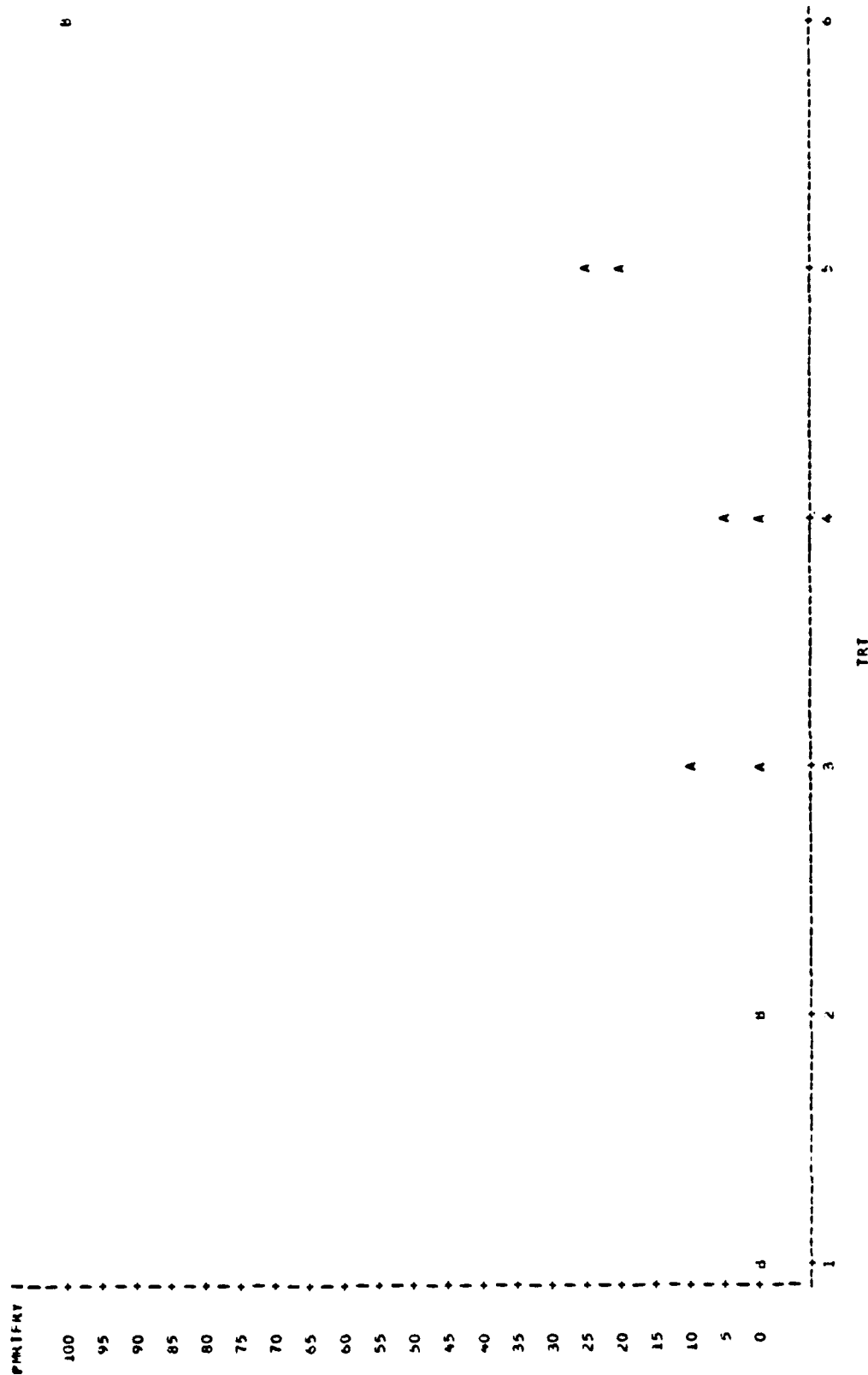


Figure VI.2 DeFoe compound C percentage fry mortality by treatment group

PLOT OF PABNEMB*TKT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

PABNEMB	1	2	3	4	5	6
99.99999						
94.44444						
88.88888						
83.33333						
77.77777						
72.22222						
66.66666						
61.11111						
55.55555						
49.99999						
44.44444						
38.88888						
33.33333						
27.77777						
22.22222						
16.66666						
11.11111						
5.55555						
0.00000						

Figure VI.3 DeFoe compound C percentage embryo abnormality by treatment group

PLOT OF PABNFRY*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

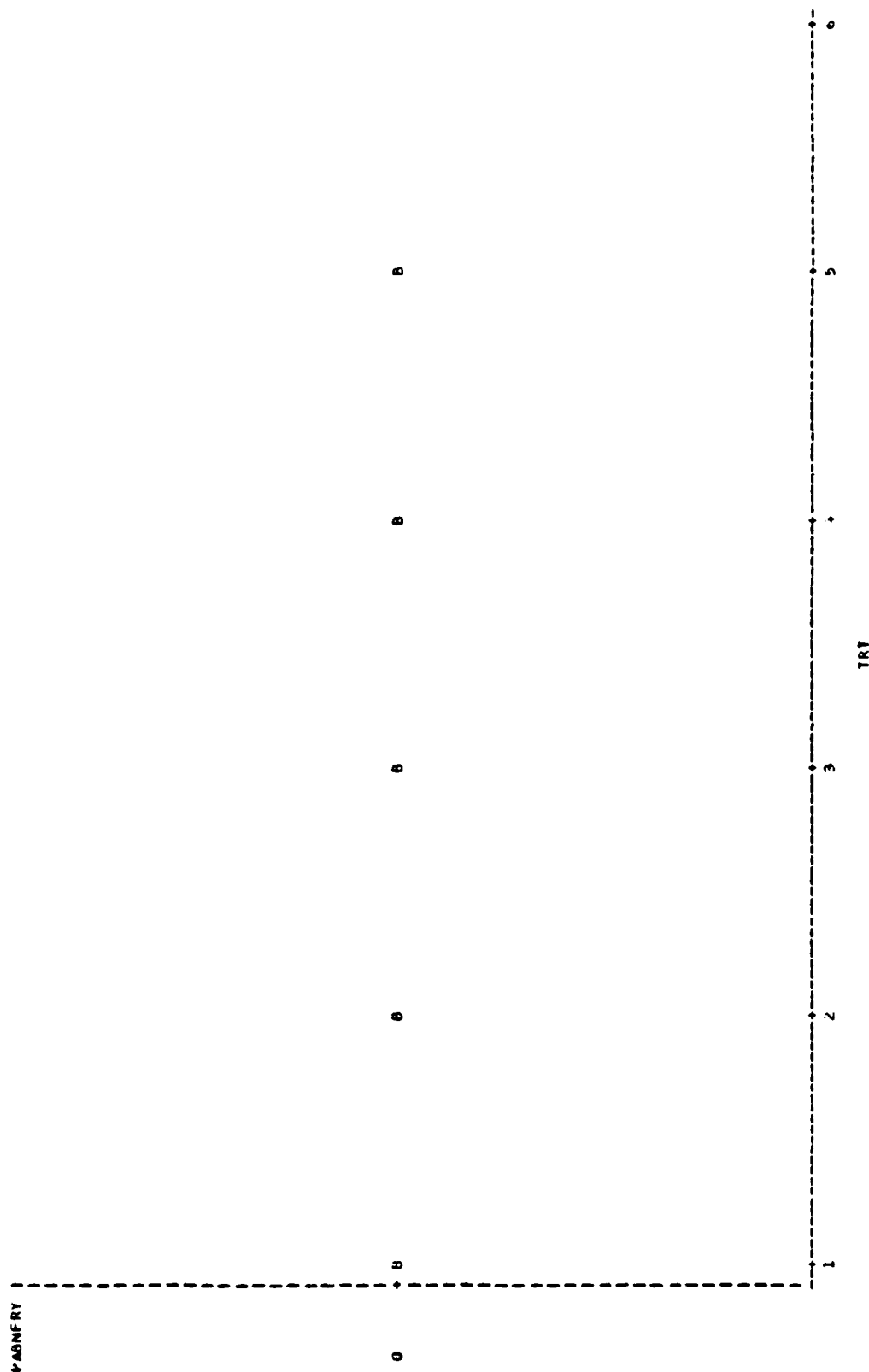


Figure VI.4 DeFoe compound C percentage fry abnormality by treatment group

PLOT OF PMRTEMB*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

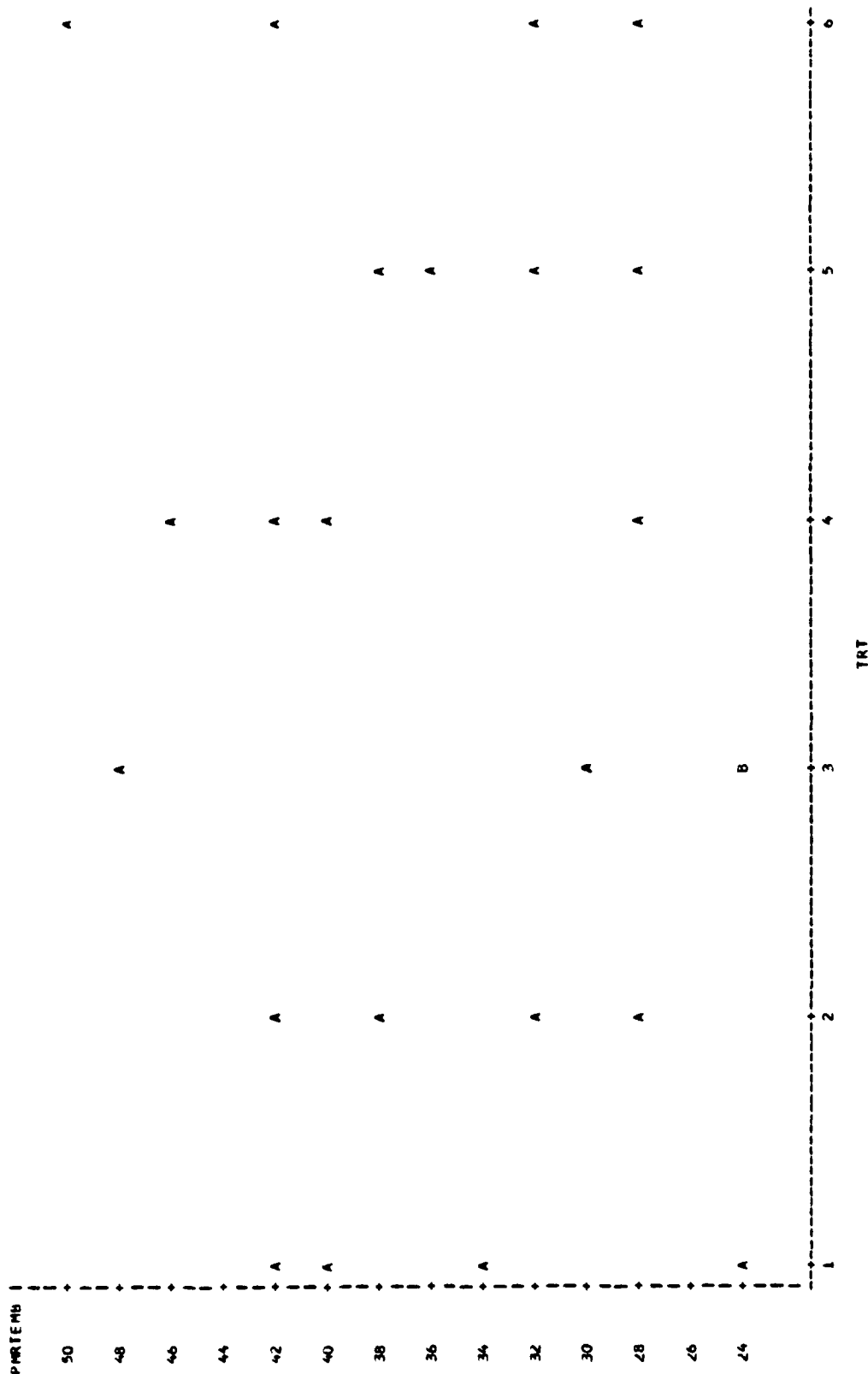


Figure VI.5 Holcombe and Phipps compound D embryo mortality by treatment group

PLOT OF PMRTFRY*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

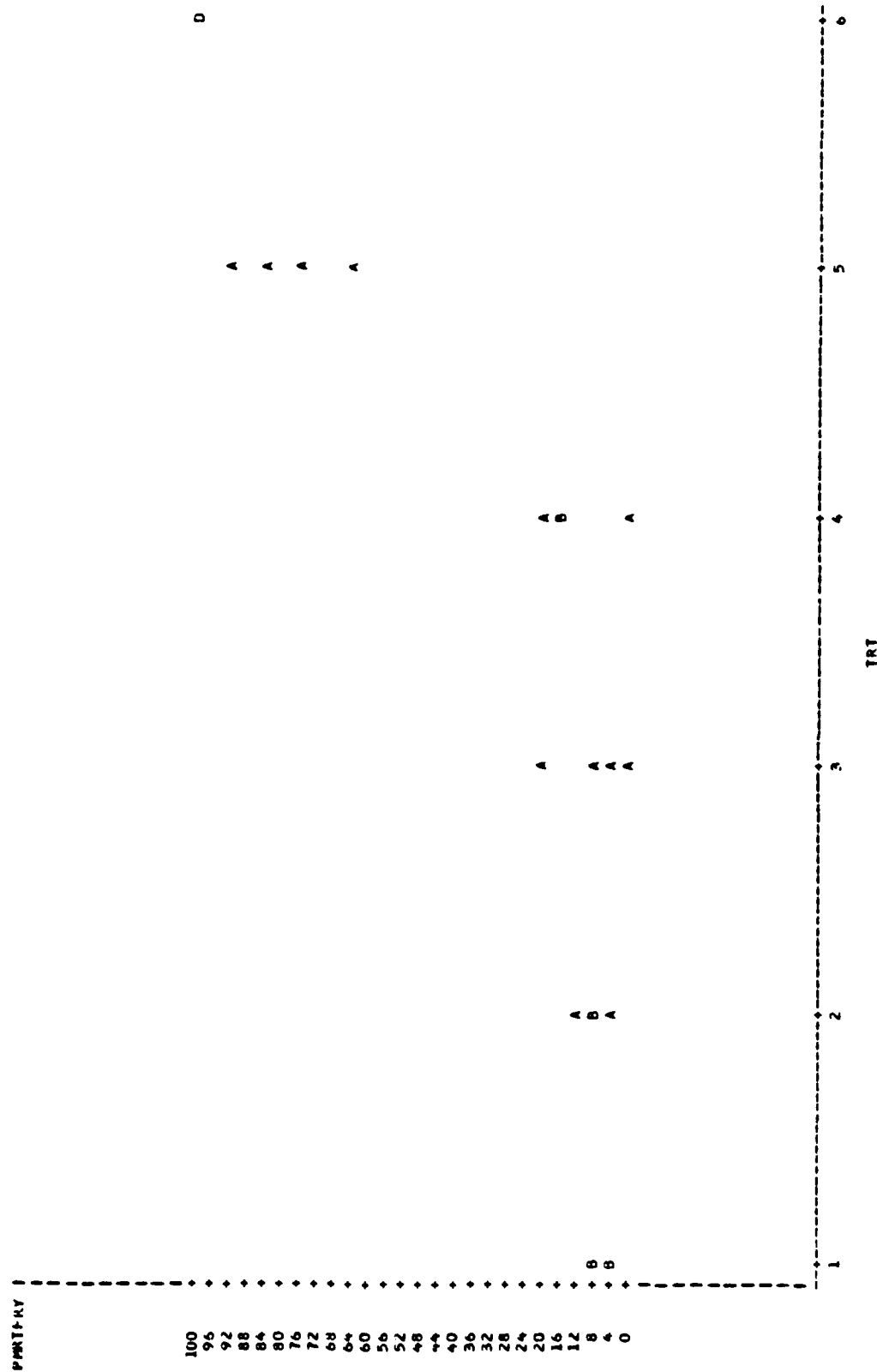


Figure VI.6 Holcombe and Phipps compound D fry mortality by treatment group

PLOT OF PABNEM#TAT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

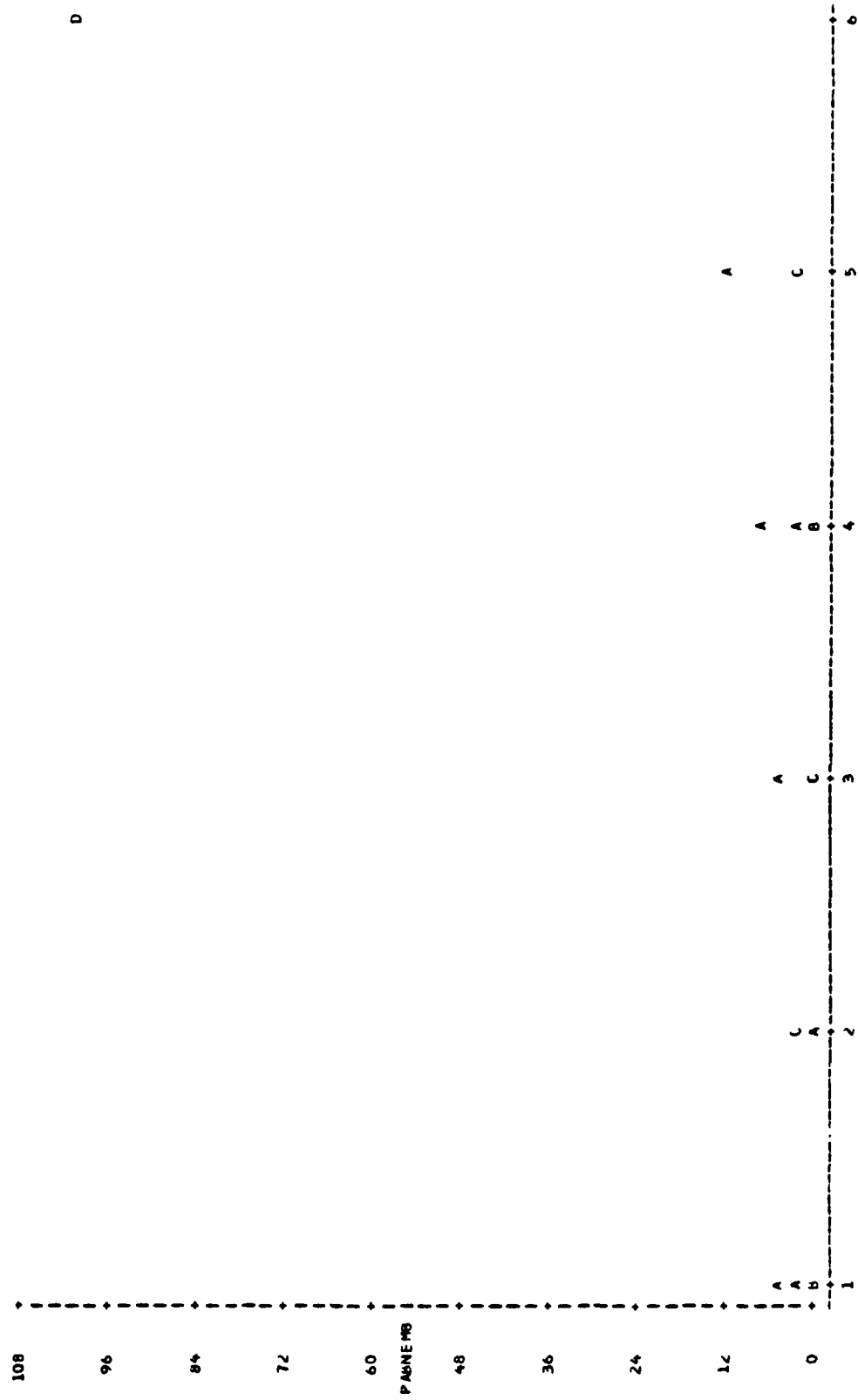


Figure VI.7 Holcombe and Phipps compound D embryo abnormality by treatment group

PLOT OF PABNFRY*TRT LEGEND: A = 1 OBS, B = 2 OBS, ETC.

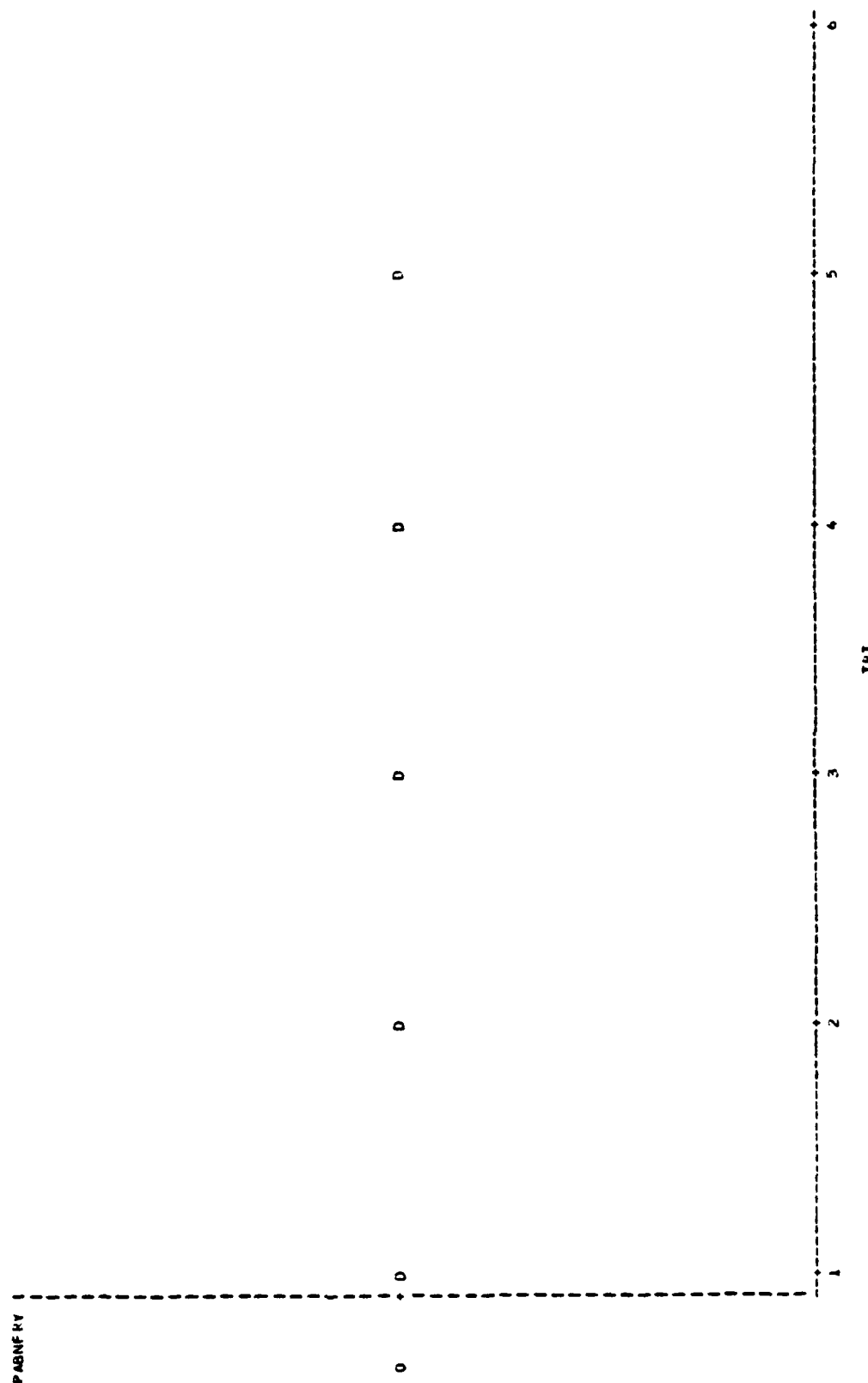


Figure VI.8 Holcombe and Phipps compound D fry abnormality by treatment group

VII. TESTING FOR TANK TO TANK HETEROGENEITY WITHIN TREATMENT GROUPS

- A. Background. In order to assess variability of response, toxicity tests generally include several fish tanks, usually two to four, within each treatment or control group. EPA guidelines for early life stage and full life cycle toxicity tests with fathead minnow [2, 8] call for at least two replicate test chambers for each treatment group. Some tests use more.

An important preliminary inference of interest in toxicology data analyses is to determine if there is any statistical evidence of variation in response among tanks within treatment groups. Such variation might be due to differences in location or handling of individual tanks, to fungus or illnesses that might invade a tank, to unforeseen accidents during the test, etc.

If evidence of tank to tank heterogeneity exists then analyses should be carried out on a per tank basis. If no evidence of tank heterogeneity exists then data might be pooled across tanks within groups and analyses carried out on a per fish basis, ignoring the replicate tanks. For example, mortality rates could be compared based on binomial theory. Such per fish analyses would provide many more degrees of freedom to estimate random error than would per tank analyses and so are more sensitive. For example if there are four tanks per group and 25 fish per tank then a per fish analysis might be based on 99 degrees of freedom per group whereas a per tank analysis would be based on just three degrees of freedom per group.

However the validity of per fish analyses rests on the absence of tank to tank heterogeneity. If there is in fact variation in response rate across tanks within treatment groups then variability estimates based on per fish analyses will underestimate the true variability of the estimates and test statistics. This will result in standard error estimates that are too small, confidence intervals that are too short, and hypothesis testing procedures that falsely reject the null hypothesis more often than their nominal rates (i.e. inflated alpha levels). It is thus important to test for the presence of tank to tank heterogeneity within treatment groups before proceeding on to the analysis of primary interest.

- B. Remarks on Some "Standard" Procedures.

Finney [11], section 9.1, pp 175 ff. suggests the following procedure for testing tank to tank heterogeneity. Fit a probit curve to the data based on pooled data across tanks within groups. Fit a probit curve to the data using the individual tanks within groups. The point estimates of the two probit fits will be exactly the same. However the residual chi squares and their

respective degrees of freedom will differ. The differences between these two residual chi squares can be interpreted as the chi square for heterogeneity among tanks within treatment groups. Similar considerations hold for the usual chi square test for homogeneity.

We have carried out this procedure using the probit fit and using the usual chi square test for homogeneity. We compare the results of these two tests. The theoretical bases of these heterogeneity tests are discussed in Appendix AVII.

We illustrate these two heterogeneity test procedures on the fry mortality data. First consider the test of compound C by DeFoe.

Probit Fit

We fitted the probit model to the treatment groups using the (natural) logarithm of concentration and excluding the control group. (Note that the same fit was obtained when the control group was included). The probit fits were carried out using the PROC PROBIT procedure in the SAS statistical computing system [12]. The data consist of I=5 treatment groups, J=2 tanks per group. Thus there are 10 responses to which we fit the two parameter probit model.

$$p_i = \Phi(\alpha - 5 + \beta \log C_i) \quad i = 2, \dots, 6$$

Figures VII.1, VII.2 contain the results of the probit fits to the individual responses and to the responses pooled across tanks within groups respectively. The analysis of variance table, as suggested by Finney, appears in the bottom portion of Figure VII.1. The upper 0.005 point of the chi square distribution with 5d.f. is 16.75. Thus this test for tank to tank heterogeneity within groups is "highly statistically significant". At face value this suggests strong statistical evidence of variation in response rate across tanks within treatment groups.

Chi Square Fit

We now carry out a chi square test of heterogeneity in response rates across tanks within groups based on the usual chi square test of homogeneity across groups. Figures VII.3, VII.4 contain the results of the chi square tests based on the individual responses and on responses pooled across tanks within groups respectively. Control group responses are included in these tests. The tests were carried out using the PROC FREQ procedure in the SAS statistical computing system [12]. The analysis of variance table suggested by Finney appears below.

Source	d.f.	SS
Lack of fit of pooled tanks about model (homogeneity)	5	182.785
Variation of individual tanks w/i tmnt groups (by subtraction)	6	0.89
Lack of fit of individual tanks about model (homogeneity)	11	183.675

Thus the heterogeneity chi square is very small. Thus there is no statistical evidence of variation among tanks within treatment groups.

Note that the conclusions arrived at from this heterogeneity chi square test are in direct contradiction to those arrived at from the heterogeneity chi square test based on the probit fit. What is the cause of the discrepancy?

There are two possible sources of difficulties. The first concerns the probability estimates in the denominators of the test statistics. In the probit based statistic the i -th group response rate in the denominator is estimated as $P_i = \hat{\phi}_i \equiv \Phi(\hat{\alpha} - 5 + \hat{\beta} \log C_i)$ whereas in the homogeneity chi square based statistic the i -th group response rate in the denominator is estimated as $\hat{P}_i = \hat{P} \equiv X_{++}/N_{++}$ for all i . The assumption of constant P values in the denominator is clearly not justified. The assumption of P_i values based on the probit model is also not good, as can be seen from the very large residual "chi square" value in Table VII.2. Thus the substantial differences in the response rate estimates that appear in the denominators of the two statistics, along with the probable inadequacies of both sets of estimates, may account for at least a portion of the discrepancy in chi square values.

The second possible source of discrepancy is based on the validity of the chi square assumption itself. The validity of the asymptotic chi square theory is dependent on the cell expected frequencies being large enough. In particular if any responses are observed in cells with very small expected frequencies then very large cell chi squares can result which can greatly inflate the statistic.

Consider the two chi square statistics for lack of fit from the probit model -- one for individual tanks and one for pooled tanks. We break out the individual components of these statistics.

Table VII.1 DeFoe Compound C Fry Mortality Data

Probit fit to groups 2, 3, 4, 5, 6 -- individual tanks
Chi square statistic for lack of fit to probit model

Grp	Tank	Mean (X _{ij})	(N _{ij})						
		Conc	Dead	Live	Total	\hat{P}_i	$N_{ij}\hat{P}_i$	$N_{ij}\hat{P}_i\hat{Q}_i$	χ^2
2	A	1.991	0	20	20	0.000015	0.000301	0.000301	0.000301
2	B	1.991	0	20	20				0.000301
3	A	5.976	0	20	20	0.002542	0.050830	0.050701	0.05096
3	B	5.976	2	18	20				74.9347
4	A	14.812	0	21	21	0.0475	0.9975	0.9501	1.0473
4	B	14.812	1	19	20		0.95	0.9409	0.00276
5	A	48.307	4	16	20	0.4227	8.454	4.8805	4.0648
5	B	48.307	5	15	20				2.4444
6	A	146.984	20	0	20	0.88356	17.6713	2.05764	2.6355
6	B	146.984	20	0	20				2.6355
									87.8165

This compares with $\chi^2 = 87.7666$ calculated from SAS PROC PROBIT

Table VII.2 Probit fit to groups 2, 3, 4, 5, 6 -- tanks pooled within groups
Chi square statistic for lack of fit to the probit model

Grp	Mean Conc	(X _{i+}) Dead	(N _{i+}) Live	(N _{i+}) Total	\hat{P}_i	$N_{i+}\hat{P}_i$	$N_{i+}\hat{P}_i\hat{Q}_i$	χ^2
2	1.991	0	40	40	0.000015	0.000602	0.000602	0.000602
3	5.976	2	38	40	0.002542	0.10166	0.10142	35.532
4	14.812	1	40	41	0.0475	1.9475	1.855	0.484
5	48.307	9	31	40	0.4227	16.908	9.761	6.407
6	146.984	40	0	40	0.88356	35.3426	4.11528	5.271
								47.695

This compares with $\chi^2 = 47.6775$ calculated by SAS PROC PROBIT

Comparison of these two chi square values clearly shows the source of the "significant" chi square for heterogeneity. Namely the tanks from group 3 have very small expected frequencies (NP) yet have observed responses. Thus these component chi square values are large and dominate the overall chi square values.

If we remove the group 3 values from the chi square statistics we have:

Separate tanks: $\chi^2 = 87.8165 - 0.05096 - 74.9347 = 12.831$

Pooled tanks: $\chi^2 = 47.686 - 35.532 = 12.154$

The relation between these two chi square statistics is then just like that of the chi square tests resulting from the contingency table tests.

Moral: Uncritical use of the chi square test for homogeneity of tanks within concentration groups recommended by Finney can lead to completely incorrect results and results contradictory to those of other homogeneity tests because of:

- small expected frequencies within cells
- response rate estimates based on particular (possibly inappropriate) model fitted

We repeated the same calculations on the fry mortality data in the test of compound D by Holcombe and Phipps.

Probit Fit

Using all six groups, a logarithmic transformation of concentration, and Abbott's correction for background response we obtain:

Source	d.f.	S.S.
Lack of fit of pooled tanks about probit model	3	0.5064
Variation of individual tanks w/i tmnt groups (by subtr)	18	21.7906
Lack of fit of individual tanks about probit model	21	22.2952

The value 21.7906 is at the upper 24 percent point of a chi square distribution with 18 d.f. and so is nonsignificant.

Chi Square Fit

There are $I = 6$ groups, $J = 4$ tanks per group. We carry out chi square tests of heterogeneity in response rates across tanks within groups based on the usual chi square test of homogeneity across groups. We obtain:

Source	d.f.	S.S.
Lack of fit of pooled tanks about model (homogeneity)	5	389.676
Variation of individual tanks w/i tmnt groups (by subtr)	18	10.198
Lack of fit of individual tanks about model (homogeneity)	23	399.874

The heterogeneity chi square is again small. There is no statistical evidence of variation among tanks within treatment groups. Note however that there is strong statistical evidence of variation in response rate from group to group, as would be expected. Thus the weights in the denominator of the heterogeneity chi square statistic are suspect.

C. Separate Heterogeneity Tests Within Treatment Groups

We have seen for DeFoe's fry mortality data that we can obtain diametrically opposite conclusions about heterogeneity of responses within treatment groups depending on whether the test for homogeneity was based on a probit model fit or on a contingency table fit. This was attributed to

1. differences in the weights used in the denominators of the chi square statistics (based on the assumed model)
2. small expected frequencies within cells that invalidate asymptotic distribution theory.

To account for problem (1), we carry out separate chi square heterogeneity tests within each concentration group, without imposing any structure on the form of the concentration-response relation. We do this by carrying out separate chi square tests within each group and then pooling the results across groups.

There is however, a technical problem associated with this approach. For many (if not most) of the responses of interest the probabilities of occurrence are fairly close to 0 or 1. Therefore the expected frequencies of occurrence can be rather small, thus invalidating the use of asymptotic chi square theory. We illustrate this phenomenon with the Holcombe and Phipps fry mortality data, broken down by group. The output (from SAS PROC FREQ) is shown in Figures VII.5 to VII.10

We see that groups 1, 2, 3, have small expected numbers of dead fry (less equal to 2.0). Group 4 has expected dead = 3.3 and group 6 has expected live per tank = 0. Thus groups 1, 2, 3, 4, 6 have small expected frequencies in at least some of the cells of the table.

We use a (relatively stringent) criterion of applicability of asymptotic chi square theory that requires that there be an expected frequency of at least 5 within each cell of the table. Only group 5¹ satisfies this criterion within the Holcombe and Phipps data.¹ We must thus base some of the within groups heterogeneity tests on exact, small sample theory.

Thus we wish to pool across groups the results of tests of homogeneity of responses among tanks within groups. Some of these tests are based on asymptotic theory while others are based on exact, small sample theory.

We have developed a computer program, EXAX2, to carry out such a procedure. We discuss this program in detail and illustrate its application in the following section.

¹Dixon and Massey [13] page 233, has a slightly more liberal criterion, namely "...none of the F_i 's (i.e. expected frequencies) is less than 1 and not more than 20% of the F_i 's are less than 5..." Again, only group 5 would satisfy this.

DEFOE COMPOUND
NOT POOLING TANKS WITHIN BORG

PROBIT ANALYSIS ON LOG (CHEAN)

Fry mortality
log concentration
not pooling tanks with control
control group deleted

ITERATION	INTERCEPT (A)	SLOPE (B)	MU	SIGMA
1	1.3304977	0.33010771	6.5307704	0.1530014
2	1.3304977	0.33010771	6.5307704	0.1530014
3	1.3304977	0.33010771	6.5307704	0.1530014
4	1.3304977	0.33010771	6.5307704	0.1530014
5	1.3304977	0.33010771	6.5307704	0.1530014
6	1.3304977	0.33010771	6.5307704	0.1530014

COVARIANCE MATRIX	INTERCEPT	SLOPE	COVARIANCE MATRIX	MU	SIGMA
INTERCEPT	4.49897031	-1.08639038	MU	0.15345155	0.02711267
SLOPE	-1.08639038	0.28235878	SIGMA	0.02711267	0.11666375

CHI-SQ = 87.7666 WITH 8 DF PROB > CHI-SQ = 0.0001

NOTE: THE LACK OF FIT OF THE POOLED TANKS MODEL IS NOT CAUSING A PROBLEM IN THIS CASE. THE LACK OF FIT OF THE INDIVIDUAL TANKS MODEL IS NOT CAUSING A PROBLEM IN THIS CASE. THE LACK OF FIT OF THE INDIVIDUAL TANKS MODEL IS NOT CAUSING A PROBLEM IN THIS CASE. (87.7666)

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS
LACK OF FIT OF POOLED TANKS MODEL	3	47.6775	

VARIATION OF INDIVIDUAL TANKS W/1 TREAT GROUPS (BY SUBTRACTION)

LACK OF FIT OF INDIVIDUAL TANKS MODEL	8	87.7666
---------------------------------------	---	---------

Figure VII.1 Probit fit to individual responses DeFoe compound C (excluding control group)

DEFEO COMPOUND C POOLING TANKS WITHIN GROUP				
PROBIT ANALYSIS ON LOG(MEAN)				
ITERATION	INTERCEPT (2)	SLOPE (2)	MU	SIGMA
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	-0.03140211	1.24728678	4.03387753	0.80174024

COVARIANCE MATRIX		COVARIANCE MATRIX	
INTERCEPT	SLOPE	MU	SIGMA
6.38675097	-1.57376025	0.22325209	0.03927580
-1.57376025	0.40902867	0.03927580	0.16980077

CHI-SQ =	47.6775	WITH	3 DF	PROB > CHI-SQ = 0.0001
----------	---------	------	------	------------------------

NOTE: THE ABOVE VARIANCES AND COVARIANCES HAVE BEEN MULTIPLIED BY THE HETEROGENEITY FACTOR $H^2 = 1.000000$.
 CHECKS FOR THE LINES OF THE COVARIANCE MATRIX (P < 0.10) IS NOT CAUSED BY SYSTEMATIC DEPARTURE FROM THE MODEL.
 A P-VALUE OF 0.10 WILL BE USED IN COMPUTING CRITICAL LIMITS.

Figure VII. 2 Probit fit to responses pooled across tanks within groups
 DeFoe compound C (excluding control group)

CONTINGENCY TABLE ANALYSIS ON DEFOL COMPOUND C DATA

TABLE OF CONTINGENCY

FREQUENCY EXPECTED CELL CHIT COL SCT	STATISTICS FOR 2-WAY TABLES												TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	
DEAD	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	0 4.3 0.00	52
LIVE	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	15.7 1.2 100.00	189
TOTAL	20	20	20	20	20	20	20	20	20	20	20	20	241

WARNING: TABLE IS SO SPARSE THAT CHI-SQUARE MAY NOT BE A VALID TEST.

CHI-SQUARE 103.675 DFF 11 PROB=0.0001

CONTINGENCY COEFFICIENT
C=0.667
CONTINGENCY COEFFICIENT

Fry mortality data
Control group included
tanks not pooled within
concentration groups

Figure VII.3 Chi square test of homogeneity based on individual responses
DeFoe compound C fry mortality.

CONTINGENCY TABLE ANALYSES ON DEFEO COMPOUND C DATA

TABLE 10. CONTINGENCY TABLE

FREQUENCY EXPECTED CELL CHI ²	TOTAL					
	1	2	3	4	5	6
DEAD	0 8.6 0.00	0 8.6 0.00	0 8.6 0.00	0 8.6 0.00	0 8.6 0.00	0 8.6 0.00
LIVE	40 21.4 100.00	40 21.4 100.00	40 21.4 100.00	40 21.4 100.00	40 21.4 100.00	40 21.4 100.00
TOTAL	40	40	40	40	40	40

STATISTICS FOR 2-WAY TABLES

CHI-SQUARE 182.785 DF= 5 PROB=0.0001

LIKELIHOOD RATIO G-SQUARE 183.426 DF= 5 PROB=0.0001

Fry mortality data
Control group included
tanks pooled within
concentration groups

Figure VII.4 Chi square test of homogeneity based on responses pooled across tanks within groups DeFoe compound C fry mortality.

HOLCOMBE AND PHIPPS COMPOUND D

TRT: 1

ASPL

FRY MORTALITY	SAS PROC FREQ	ASPL				TOTAL
		A	B	C	D	
DEAD		1.3 0.2	2.3 0.2	1.5 0.2	1.3 0.2	6
LIVE		23.3 0.0	23.3 0.0	23.5 0.0	23.5 0.0	94
TOTAL		25	25	25	25	100

FRY MORTALITY

$$\chi^2 = 0.8$$

Figure VII.5 Chi square test of homogeneity of percent fry mortality in Group 1. Data from Holcombe and Phipps compound D

HOLCOMBE AND PHIPPS COMPOUND D

TRI = 2

NEPL

FRY MORTALITY

SAS PROC FREQ

	A	B	C	D	TOTAL
DEAD	3 2.0 0.5	1 2.0 0.5	2 2.0 0.0	3 3.0 0.0	8
LIVE	22 23.0 0.0	24 23.0 0.0	23 23.0 0.0	23 23.0 0.0	92
TOTAL	25	25	25	25	100

FRY MORTALITY

$$\chi^2_3 = 1$$

Figure VII.6 Chi square test of homogeneity of percent fry mortality in group 2. Data from Holcombe and Phipps compound D

HOLCOMBE AND PHIPPS COMPOUND D

TAT = 3

REPL

FRY MORTALITY	SAS PROC FREQ	REPL				TOTAL
		A	B	C	D	
DEAD		2 2.0 0.0	1 1.0 0.0	5 5.0 0.0	1 1.0 0.0	8
LIVE		23 23.0 0.0	25 25.0 0.0	20 20.0 0.0	24 24.0 0.0	92
TOTAL		25	25	25	25	100

FRY MORTALITY

$$\chi^2_3 = 7.6$$

Figure VII.7 Chi square test of homogeneity of percent fry mortality in group 3. Data from Holcombe and Phipps compound D

HOLCOMBE AND PHIPPS COMPOUND D

TABLE 4

		REPL				TOTAL
		A	B	C	D	
FRY MORTALITY	DEAD	4 3.3 0.2 16.00	5 3.3 0.9 20.00	4 3.3 0.2 16.00	0 3.3 3.3 0.00	23
	LIVE	21.8 21.8 0.0 84.00	20 21.8 0.1 80.00	21.8 21.8 0.0 84.00	25 21.8 0.0 100.00	87
TOTAL		25	25	25	25	100
SAS PROC FREQ						

FRY MORTALITY

$$\chi^2 = 52$$

Figure VII.8 Chi square test of homogeneity of percent fry mortality in group 4. Data from Holcombe and Phipps compound D

HOLCOMBE AND PHIPPS COMPOUND D.

TATE = 5

REPL

COLL

FRY MORTALITY
X² TEST
CELL COUNTS
DEAD
LIVE

FRY MORTALITY

SAS PROC FREQ

	A	B	IC	D	TOTAL
DEAD	23 19.8 92.00	21 19.8 94.00	16 19.8 64.00	19 19.8 76.00	79
LIVE	2 5.3 8.00	4 5.3 16.00	9 5.3 36.00	6 5.3 24.00	21
TOTAL	25	25	25	25	100

FRY MORTALITY

$$\chi^2 = 6.7$$

Figure VII.9 Chi square test of homogeneity of percent fry mortality in group 5. Data from Holcombe and Phipps compound D

HOLCOMBE AND PHIPPS COMPOUND D

TAT= 6

REPL

FRY MORTALITY	REPL				TOTAL
	A	B	C	D	
DEAD	25 25.0 100.00	25 25.0 100.00	25 25.0 100.00	25 25.0 100.00	100
LIVE	0 : :	0 : :	0 : :	0 : :	0
TOTAL	25	25	25	25	100

FRY MORTALITY

SAS PROC FREQ

FRY MORTALITY

$$\chi^2 = 0$$

Figure VII.10 Chi square test of homogeneity of percent fry mortality in group 6. Data from Holcombe and Phipps compound D

VIII. EXAX2 -- A COMPUTER PROGRAM TO TEST FOR HETEROGENEITY OF RESPONSES AMONG TANKS WITHIN GROUPS

We saw in the previous section for DeFoe's fry mortality responses from the test on compound C that we obtained diametrically opposite impressions about the existence of tank to tank heterogeneity within groups depending on whether we based our homogeneity test on a probit model fit or on a contingency table fit. This was attributed to

1. Differences in weights used in the denominators of the chi square statistics (based on the assumed model)
2. Small expected frequencies that produce substantial departures from the asymptotic distribution theory.

To take account of problem (1), we use the strategy of carrying out separate chi square heterogeneity tests within each concentration group, without imposing any structure on the concentration response relation.

There is, however, a technical problem associated with this scheme. For many (if not most) of the responses of interest the response probabilities of occurrence are fairly close to 0 or 1. Therefore the expected frequencies of occurrence can be rather small, thus invalidating the asymptotic chi square theory, upon which most of the standard tests are based. We saw this in connection with the fry mortality data from the Holcombe and Phipps test on compound D.

We have developed a computer program, EXAX2, that overcomes this problem. It carries out separate chi square tests within each treatment group, based on asymptotic theory when the expected frequencies within cells are large enough and based on exact, small sample theory when the expected frequencies within cells are small. Thus heterogeneity tests using EXAX2 are applicable even with the relatively small sample sizes and relatively extreme response rates encountered in fish toxicity tests. The theory underlying the program and instructions for its use are described in [14] which included as Appendix AVIII.2. In the body of the section we describe the basis of the calculations in EXAX2 and illustrate its application with examples.

EXAX2 pools the results of tests for heterogeneity in each of $I(I > 1)$ $2 \times K$ independent contingency tables (representing I groups, K tanks per group). The homogeneity test within each group is based on the usual chi square statistic, using either its asymptotic distribution or its exact small sample distribution, as appropriate. The following approach is used.

1. Within each concentration group, the chi square for homogeneity among the K tanks is calculated. Let X_{ij} , N_{ij} represent the

number of dead fish and the total number of fish respectively in the j -th tank of the i -th group. Let \hat{p}_i be a pooled estimate of response probability in the i -th group and $\hat{q}_i = 1 - \hat{p}_i$. Then the chi square for group i is

$$\chi_i^2 = \sum_{j=1}^K \frac{(X_{ij} - N_{ij}\hat{p}_i)^2}{N_{ij}\hat{p}_i\hat{q}_i}$$

If $\hat{p}_i = 0$ or $\hat{q}_i = 0$ (corresponding to zero percent or 100 percent observed mortality) the table is degenerate and $\chi_i^2 = 0$ by definition.

2. If all the expected frequencies ($N_{ij}\hat{p}_i, N_{ij}\hat{q}_i$) are greater than a specified cutoff value (we currently use five), asymptotic theory is used. Thus the observed significance level of χ_i^2 is based on the chi square distribution with $K - 1$ d.f.
3. If one or more of the expected frequencies is less than the cutoff level, then the exact distribution of χ_i^2 , conditional on the observed marginal totals, is used. The observed significance level is based on this exact distribution. This approach is described in Agresti and Wackerly [15]. The exact distribution of χ_i^2 is computed by systematically enumerating all possible tables having the given margins using the algorithm in Boulton and Wallace [16] and the associated probabilities due to March [18].
4. Let A_i denote the observed significance level in the i -th group. We pool the A_i 's over groups to obtain an overall test by an approach analogous to Fisher's method as described in Littell and Folks [19, 20]. For each group we calculate, based on exact or asymptotic theory, $-2\ln A_i$ and its mean and variance under the null hypothesis of homogeneity.
5. The observed significance levels are pooled into a single statistic by calculating

$$Z = \frac{\left[\sum_{i=1}^I (-2\ln A_i) \right]^{1/2} - \left[\sum_{i=1}^I E(-2\ln A_i) \right]^{1/2}}{\left[\sum_{i=1}^I \text{Var}(-2\ln A_i)/4 \sum_{i=1}^I E(-2\ln A_i) \right]^{1/2}}$$

Z is referred to a standard normal distribution. The null hypothesis of tank to tank homogeneity is rejected for large values of Z . (The square root transformation is used because it represents the variance stabilizing transformation, under asymptotic

theory, for $\sum_i (-2\ln A_i)$ and thus probably improves the normality approximation.)

In addition to calculating preliminary tests of tank to tank heterogeneity within treatment groups, EXAX2 can carry out several other statistical procedures useful in the analysis of data from aquatic toxicity tests. In particular it can:

- Pool data across tanks within groups and test for heterogeneity of response rate across groups by use of the chi square test and either exact small sample theory or asymptotic large sample theory.
- Calculate confidence intervals on the odds ratios of treatment groups to control group using the exact non-central distribution of Fisher's exact test statistic.

These applications will be discussed in detail in later section.

We now consider several illustrations of the use of EXAX2 for tests of heterogeneity among tanks within groups. The EXAX2 outputs are shown in the referenced figures. The observed and expected cell frequencies are indicated. If any of the expected cell frequencies are lower than the (user-specified) cutoff of 5, exact distribution theory is used. The exact distribution of chi square, conditional on the marginal totals, is enumerated and displayed. The observed value of chi square, the observed significance level, $-2\ln A_i$, $E(-2\ln A_i)$, $\text{Var}(-2\ln A_i)$ are calculated. The six independent tests are combined by summing $-2\ln A_i$, $E(-2\ln A_i)$, $\text{Var}(-2\ln A_i)$ over groups and calculating Z.

DeFoe compound C

- a) Embryo mortality
- b) Fry mortality

Holcombe and Phipps, Compound D

- a) Embryo mortality
- b) Fry mortality

Jarvinen, compound B

- a) Embryo mortality
- b) Fry mortality

DeFoe Compound C

Embryo Mortality

There are two tanks per treatment group, 50 embryo per tank. The results from the EXAX2 calculations are shown in Figures VIII.1 to VIII.6 and are summarized below*. The pooled significance level calculations (using Fisher's method) are presented below the results for group 6. The probability of a standard normal deviate exceeding 0.865 is 0.19.

Embryo Mortality

Trt	Method	(Chi sq) XSQØBS	(A _i) AI	(-2lnA _i) YY	E(-2lnA _i) EY	Var(-2lnA _i) VARY
1	asympt	0.37205	0.54189	1.22540	2.00	4.00
2	asympt	6.76271	0.00931	9.35371	2.00	4.00
3	asympt	0.38610	0.53436	1.25338	2.00	4.00
4	asympt	1.07250	0.30038	2.40541	2.00	4.00
5	asympt	1.05086	0.30531	2.37286	2.00	4.00
6	asympt	0.0	1.00	0.0	2.00	4.00

SY = 16.611

YMU = 12.00

SVARY = 24.00

Z = 0.86483

Except for group 2, where the response from tank 1 appears to be an outlier, there is no statistical evidence of tank to tank heterogeneity within groups.

Fry Mortality

There are two tanks per treatment group, 20 fry per tank. The result from the EXAX2 calculations are shown in Figures VIII.7 to VIII.12 and are summarized below. The pooled significance level calculations are presented below the results for group 6. The probability of a standard normal deviate exceeding 0.175 is 0.43.

Fry Mortality

Trt	Method	(Chi sq) XSQØBS	(A _i) AI	(-2lnA _i) YY	E(-2lnA _i) EY	Var(-2lnA _i) VARY
1	(Row total = 0)			0.0	0.0	0.0
2	(Row total = 0)			0.0	0.0	0.0
3	EXACT	2.10526	0.48718	1.43825	0.70068	0.5168
4	EXACT	1.07625	0.48780	1.43568	0.70033	0.51499
5	EXACT	0.14337	1.00	0.0	1.12060	2.7544
6	(Row total = 0)			0.0	0.0	0.0

SY = 2.8739

YMU = 2.5216

SVARY = 3.7862

Z = 0.17516

*Figures VIII.1 to VIII.36 are contained in Appendix AVIII.1.

Fry Mortality

Trt	Method	(Chi Sq) XSQOBS	(A _i) AI	(-2lnA _i) YY	E(-2lnA _i) EY	Var(-2lnA _i) VARY
1	EXACT	0.70922	1.00	0.0	1.3544	2.7875
2	EXACT	1.08696	0.95647	0.08901	1.543	3.4415
3	EXACT	7.06870	0.07579	5.15948	1.543	3.4415
4	EXACT	5.21662	0.18667	3.35688	1.7349	2.6708
5	asympt	6.44967	0.01967	4.77915	2.0	4.0
6	(Row total = 0)		1.0000	0.0	0.0	0.0

SY = 13.385

YMU = 8.1752

SVARY = 17.341

Z = 1.09755

Groups 1, 2, 4, 6 show no statistical evidence of tank to tank heterogeneity. Groups 3, 6 show some marginal suggestion of tank to tank heterogeneity. It is interesting to note that in direct analogy with the results for embryo mortality, tank 3 of group 3 has about twice the mortality of the other tanks in the group. This "coincidence" should be further investigated to determine if this increased mortality has a systematic cause. Overall, Z = 1.10. The probability of a standard normal random variable exceeding this value by chance is 0.136. Thus there is at most a marginal suggestion of some possible tank to tank variation, but nothing conclusive.

Jarvinen Compound B

Embryo Mortality

There are two tanks per treatment group, approximately 50 embryos per tank (actually, between 48 and 57 with an average of 51.2). The results from the EXAX2 calculations are shown in Figures VIII.25 to VIII.30 and are summarized below. The pooled significance level calculations are given below the results for group 6. The probability of a standard normal deviate exceeding 2.54 is 0.005.

Embryo Mortality

Trt	Method	(Chi Sq) XSQOBS	(A _i) AI	(-2lnA _i) YY	E(-2lnA _i) EY	Var(-2lnA _i) VARY
1	Asympt	6.51208	0.01071	9.07234	2.0	4.0
2	EXACT	1.78430	0.27477	2.58361	1.52817	3.23307
3	EXACT	3.05250	0.15951	3.67136	1.17120	2.77808
4	EXACT	0.00085	1.0000	0.00000	1.04511	1.38125
5	EXACT	0.74812	0.43704	1.65548	1.45242	2.94750
6	EXACT	4.75938	0.05966	5.63828	1.52125	3.17408

SY = 22.62107

YMU = 8.71816

SVARY = 17.51398

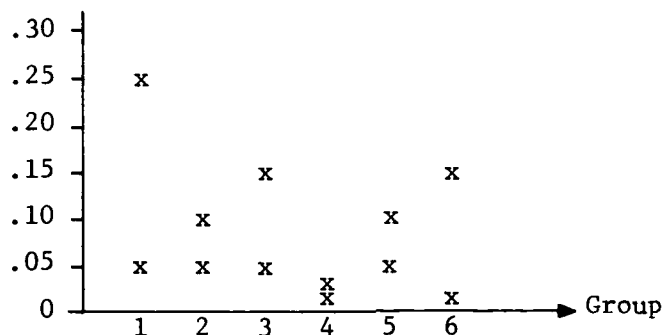
Z = 2.54488

Groups 1 and 6 show significant differences between mortality rates in replicate tanks. Overall ($Z = 2.54$) the heterogeneity statistic is significant at $\alpha \approx 0.005$ level. Thus overall there is strong statistical evidence of tank to tank heterogeneity.

Group 1 shows considerable tank to tank heterogeneity in response, group 6 shows moderate tank to tank heterogeneity in response, and group 3 shows marginal tank to tank heterogeneity in response.

If we plot mortality rate by group number we obtain

Embryo mortality rate



In agreement with the DeFoe and Holcombe and Phipps embryo mortality results, we see no trend in embryo mortality rate with increasing toxicant concentration. We see tank to tank heterogeneity in group 1 and to a lesser extent in groups 3, 6. There is the suggestion that the response from tank 1 of group 1 might be an outlier. This will be considered further in section X.

Fry Mortality

There are two tanks per treatment group, approximately 15 fry per tank (between 14 and 16 with an average of 14.9). The results from the EXAX2 calculations are shown in Figures VIII.31 to VIII.36 and are summarized below. The pooled significance level calculations are given below the results for group 6. The probability of a standard normal deviate exceeding -0.84 is 0.80.

Fry Mortality

Trt	Method	(chi sq) XSQOBS	(A ₁) AI	(-2lnA ₁) YY	E(-2lnA ₁) EY	Var(2lnA ₁) VARY
1	(Row total = 0)			0.0	0.0	0.0
2	(Row total = 0)			0.0	0.0	0.0
3	(Row total = 0)			0.0	0.0	0.0
4	(Row total = 0)			0.0	0.0	0.0
5	asympt	0.13393	0.71439	0.67264	2.0	4.0
6	(Row total = 0)			0.0	0.0	0.0

SY = 0.67264
Z = -0.84013

YMU = 2.0

SVARY = 4.0

This test does not reveal the concentration-response curve very well. In groups 1-4 no fry died while in group 6 all the fry died. Thus the tables are degenerate in 5 of the 6 treatment groups. In group 5 there is no suggestion of tank to tank heterogeneity. Thus overall, there is no suggestion of tank to tank heterogeneity.

In summary we have seen several different degrees of tank to tank heterogeneity within groups in the three toxicity tests studied. With respect to embryo mortality the DeFoe test shows no suggestion of heterogeneity with the exception of an isolated outlier, the Holcombe and Phipps test reveals possible suggestion of heterogeneity, and the Jarvinen test reveals strong suggestion of heterogeneity but this may also be due to an outlier. With respect to fry mortality the DeFoe and Jarvinen tests show no suggestion of tank to tank heterogeneity within groups. The Holcombe and Phipps tests shows possible suggestion of tank to tank heterogeneity but the aberrant looking response originates in precisely the same tank as does the aberrant looking embryo mortality response. This raises questions about both responses. In brief, there does not appear to be very much tank to tank heterogeneity within groups and that which does occur may be due to isolated outlying results.

IX. ADJUSTMENTS TO ACCOUNT FOR TANK TO TANK HETEROGENEITY WITHIN TREATMENT GROUPS

A. Background, Derivations, and Discussions

We have considered the problem of testing for tank to tank heterogeneity within treatment groups. The results of such tests will influence the way we treat the data in subsequent analyses. Many methods for analyzing qualitative dose response data tacitly assume that there is no tank to tank variation in response rates within groups. Binomial distribution theory is used on data pooled across tanks within groups. Sometimes the assumption of lack of tank to tank heterogeneity is reasonable, as we have seen with the DeFoe and Jarvinen fry mortality data. Sometimes there is borderline statistical evidence of tank to tank heterogeneity, as was the case with the Holcombe and Phipps test (both for the embryo mortality and fry mortality data). In other situations, such as in the embryo mortality data from Jarvinen's test on methyl parathion encapsulated, there is stronger statistical evidence of tank to tank heterogeneity, in at least some of the treatment groups.

In this section we consider methods for accounting for tank to tank heterogeneity when it exists. Three main approaches are possible.

1. We can formulate models that explicitly account for tank to tank heterogeneity within groups and fit these models to the data by specialized techniques such as maximum likelihood estimation, using special purpose computer programs. Two such models are the beta binomial [21] and the correlated binomial [22]. This approach requires the formulation of specialized models and development of specialized programs to implement these analyses. Thus such analyses will be difficult for experimenters to carry out and the results of such analyses will be more difficult to interpret.
2. We can carry out analyses on a per tank basis rather than on a per fish basis. That is, summary values of such as percent mortality, average weight gain, etc are calculated within each tank and are then used as basic values for subsequent analyses. This is currently the most commonly used approach for analyzing fish toxicity data. While it does correctly account for possible tank to tank heterogeneity, it does so at the cost of considerable reduction in sensitivity. Namely, the data from perhaps 50 to 100 fish or embryos per group are summarized by just two to four summary values. This leaves very few degrees of freedom for estimating error and so diminishes the sensitivity of the subsequent procedures.

3. We can adjust the data to reflect the increased variability due to tank to tank heterogeneity and then use "standard", binomial based techniques on the adjusted "data". This third approach is a workman-like approach and has the dual virtues of being simple to carry out and of permitting the use of "standard" statistical procedures and computer programs for subsequent analyses.

Heterogeneity among tanks within groups can be alternatively regarded as correlation among the responses of the various fish within the same tank. Such correlation is usually positive and this has the effect of increasing the variability of statistics over and above that which would be assumed under a binomial model.

The increased variability can be accounted for by reducing the actual sample size in each tank to an effective sample size and then disregarding the correlation. The number of responses is reduced proportionately so that the observed response rate within each tank remains constant. Suppose for instance there are 40 embryo in a tank and 8 die. We thus have an observed response rate of 0.20. Suppose that the responses within each tank are positively correlated and the variance of \hat{p} is inflated 20 percent by this correlation. That is

$$\text{Var}(\hat{p}) = \frac{1.2p(1-p)}{40}$$

Then we can regard the effective sample size within that tank as $40/1.2 = 33.33$. To maintain the response rate at the observed level of .20 we adjust the number of responses down to a corresponding effective number $8/1.2 = 6.67$. We then analyze the data from this tank by ignoring the correlation and treating the data as if we have 6.67 responses in 33.33 trials. All the standard analysis procedures, predicated on the assumption of no tank to tank heterogeneity within groups, can be applied to the modified "data".

The per tank analyses mentioned in paragraph 2 can be regarded as a limiting case of data adjustment where we adjust the effective sample sizes within tanks all the way down to 1.

We now consider the calculation of adjustment factors. Motivation for the adjustment procedure comes from the form of the beta binomial model [21]. Namely suppose X_{ij} is the number of responses within tank j of treatment group i . The beta binomial model extends the binomial model to allow for tank to tank variation within groups. Thus we assume

$$X_{ij} \sim \text{Binomial}(N_{ij}, p_{ij}) \quad \text{conditional on } p_{ij}, \quad \begin{matrix} j = 1, \dots, J \\ i = 1, \dots, I \end{matrix}$$

where $p_{ij} \sim \text{Beta}(\alpha_i, \beta_i)$

and N_{ij} are fixed.

$$\text{Let } \mu_i \equiv \frac{\alpha_i}{\alpha_i + \beta_i}$$

$$\theta_i \equiv \frac{1}{\alpha_i + \beta_i}$$

$$\text{Then } E(p_{ij}) = \mu_i$$

$$\text{Var}(p_{ij}) = \mu_i(1 - \mu_i) \frac{\theta_i}{1 + \theta_i}$$

When $\theta_i = 0$, $\text{Var}(p_{ij}) = 0$ and we are back to the case of no tank to tank heterogeneity, at least within the i -th group. The larger θ_i is, the greater is the extent of tank to tank heterogeneity. θ_i varies between 0 and ∞ .

Now consider the distribution of X_{ij} .

$$L(X_{ij}, p_{ij}) \sim \text{Binomial}(N_{ij}, p_{ij})$$

$$L(p_{ij}) \sim \text{Beta}(\alpha_i, \beta_i).$$

These two facts imply that X_{ij} has the marginal beta binomial distribution with probability function

$$P(X_{ij} = x) = \binom{N_{ij}}{x} \frac{B\left(\frac{\mu_i}{\theta_i} + x, \frac{1 - \mu_i}{\theta_i} + N_{ij} - x\right)}{B\left(\frac{\mu_i}{\theta_i}, \frac{1 - \mu_i}{\theta_i}\right)} \quad x = 0, 1, \dots, N_{ij}$$

It can be shown directly that

$$E(X_{ij}) = N_{ij} \mu_i$$

$$\text{Var}(X_{ij}) = N_{ij} \mu_i (1 - \mu_i) \left[\frac{1 + N_{ij} \theta_i}{1 + \theta_i} \right] \quad 0 < \theta_i < \infty$$

We see that the variance of X_{ij} is inflated over and above a binomial variance by a multiplicative factor.

Suppose $N_{ij} = N_i$ $j = 1, \dots, J$. This assumption is reasonable in fish toxicity tests where N_{ij} represents the number of fish or embryos within the j -th tank of the i -th group. In this case the multiplicative factor becomes $[(1 + N_i \theta_i)/(1 + \theta_i)] \equiv K_i$, $j = 1, \dots, J$. Thus $\text{Var}(X_{ij}) = N_i K_i \mu_i (1 - \mu_i)$ where $1 \leq K_i < \infty$. Define $\hat{p}_{ij} = X_{ij}/N_{ij}$. \hat{p}_{ij} is the observed response proportion.

Therefore

$$\text{Var}(\hat{p}_{ij}) = \frac{K_i}{N_i} \mu_i (1 - \mu_i) \quad j = 1, \dots, J$$

Thus the effective sample size is N_i/K_i . As the extent of tank to tank heterogeneity approaches 0 (i.e. $\theta_i \rightarrow 0$), K_i approaches 1 and N_i/K_i approaches N_i . As the extent of tank to tank heterogeneity gets greater and greater, $K_i \rightarrow N_i$ and so N_i/K_i approaches 1. Thus the two extreme situations are not adjusting the within tank sample sizes at all and adjusting the within tank sample size down to 1. The latter adjustment resembles performing analyses on a per tank basis rather than on a per fish basis. Thus method 2 for accounting for tank to tank heterogeneity can be regarded as an extreme case of method 3. Note that if $N_i \equiv N$ and $\theta_j \equiv \theta$ for all i , then $K_i \equiv K$ for all i .

The procedure suggested here for calculating adjustment factors is motivated by the results based on beta binomial theory, but is simpler to carry out.

Let X_{ij} , N_i denote the number of responses and the total number of fish respectively within tank j of group i , $j = 1, \dots, J$. Let $\hat{p}_{ij} \equiv X_{ij}/N_i$. Let p_i denote the average response rate within the i -th group. The actual variance of \hat{p}_{ij} is the binomial theory variance multiplied by the inflation factor K_i . Thus

$$K_i = \frac{\text{Var}(\hat{p}_{ij})}{[p_i(1 - p_i)/N_i]}$$

This suggests that we can estimate K_i by estimating $\text{Var}(\hat{p}_{ij})$ and p_i by their sample analogues. Let

$\bar{N}_i \equiv J^{-1} \sum_{j=1}^J N_{ij}$, $\hat{p}_i \equiv J^{-1} \sum_{j=1}^J \hat{p}_{ij}$, $\hat{\text{Var}}(\hat{p}_{ij}) \equiv (J - 1)^{-1} \sum_{j=1}^J (\hat{p}_{ij} - \hat{p}_i)^2$ denote the average sample size, the average observed response rate, and the sample variance of response rates within the i -th group respectively. Note that the N_{ij} 's are generally nearly equal in fish toxicity data. We estimate K_i as

$$\hat{K}_i = \hat{\text{Var}}(\hat{p}_{ij}) / [\hat{p}_i(1 - \hat{p}_i) / \bar{N}_i]$$

The numerator of this ratio is the observed variance among the \hat{p}_{ij} 's while the denominator is the variance that would be expected just due to binomial variation. We adjust each X_{ij} , N_{ij} in the group downward by a factor \hat{K}_i .

Notes:

1. K_i is necessarily greater than 1 but \hat{K}_i may not be. If $\hat{K}_i < 1$ or if there is no statistical evidence of tank to tank heterogeneity then we should not adjust sample sizes.
2. Assuming binomial theory when there is in fact tank to tank heterogeneity results in underestimation of the variabilities of

the various statistics calculated. Thus hypothesis tests comparing treatment group and control group response rates will reject more often than they should, thereby resulting in underestimation of no-effect levels. However the opposite effect occurs with respect to inferences about safe concentrations based on dose response curves. Underestimation of variability results in overly large lower confidence bounds on safe concentration. A nominal 95% lower confidence bound may in fact be just an 80% lower confidence bound.

3. The decision as to when to adjust sample sizes downward should be reasonably liberal, perhaps when there is statistical evidence of tank to tank heterogeneity at the $\alpha = .20$ or the $\alpha = .25$ level. However \hat{K}_i must always be greater than or equal to 1.
4. The calculation of \hat{K}_i by means of ratios of variances is inefficient. A more precise way of determining K_i would be to estimate θ_i from the data by maximum likelihood estimation and substitute this estimate, $\hat{\theta}_i$, into the expression for K_i . Such an estimate would always be greater than or equal to 1. However such an approach would require special purpose programs. The estimation of \hat{K}_i as discussed in this section is simpler and can be carried out by hand calculation. However in the future we will look in calculation of \hat{K}_i 's by means of maximum likelihood estimation based on the beta binomial model.
5. We can calculate separate K_i 's for each treatment group based on the responses solely from the tanks in that group. Alternatively if $N_i = N$ for all i and if $\theta_i = \theta$ for all i then $K_i = K$ for all i . We can then calculate a common inflation factor \hat{K} for all treatment groups. The question of whether we should fit separate adjustment factors within each group or a single common factor is a research problem in its own right. We defer the answer to that question to future work and in this report confine attention to fitting common adjustment factors for all groups.

If it is sensible on biological grounds, results on tank to tank heterogeneity observed in previous similar tests might be combined with current results to obtain a more accurate adjustment factor.

6. The adjustment procedure might take into account the statistical precisions of the estimates \hat{K}_i , \hat{K} . A conservative way to do this would be to use upper confidence bounds on \hat{K}_i , \hat{K} as adjustment factors rather than the point estimates. This modification will also await future work.

B. Illustrations

We illustrate the application of this adjustment procedure to several sets of data. First consider the fry mortality data of Holcombe and Phipps for compound D. From the preliminary test of tank to tank heterogeneity within treatment groups we conclude that there is at most marginal

statistical evidence of tank to tank heterogeneity within treatment groups. (The observed significance level is 0.14). In this example $J = 4$, $N_{ij} = 25$ for all i, j .

Group 1: $\hat{p}_{11} = 0.08, \hat{p}_{12} = 0.08, \hat{p}_{13} = 0.04, \hat{p}_{14} = 0.04, \hat{p}_1 = 0.06,$
 $\bar{N}_1 = 25, \hat{\text{Var}}(\hat{p}_{1j}) = 0.00053, \hat{p}_1(1 - \hat{p}_1)/\bar{N}_1 = 0.00226$

$$\hat{K}_1 = \frac{\hat{\text{Var}}(\hat{p}_{1j})}{[\hat{p}_1(1 - \hat{p}_1)/\bar{N}_1]} = \frac{0.00053}{0.00226} = 0.236$$

Group 2: $\hat{p}_{21} = 0.12, \hat{p}_{22} = 0.04, \hat{p}_{23} = 0.08, \hat{p}_{24} = 0.08, \hat{p}_2 = 0.08,$
 $\bar{N}_2 = 25, \hat{\text{Var}}(\hat{p}_{2j}) = 0.00107, \hat{p}_2(1 - \hat{p}_2)/\bar{N}_2 = 0.00294$

$$\hat{K}_2 = \frac{\hat{\text{Var}}(\hat{p}_{2j})}{[\hat{p}_2(1 - \hat{p}_2)/\bar{N}_2]} = \frac{0.00107}{0.00294} = 0.362$$

Group 3: $\hat{p}_{31} = 0.08, \hat{p}_{32} = 0.00, \hat{p}_{33} = 0.20, \hat{p}_{34} = 0.04, \hat{p}_3 = 0.08,$
 $\bar{N}_3 = 25, \hat{\text{Var}}(\hat{p}_{3j}) = 0.00747, \hat{p}_3(1 - \hat{p}_3)/\bar{N}_3 = 0.00294$

$$\hat{K}_3 = \frac{\hat{\text{Var}}(\hat{p}_{3j})}{[\hat{p}_3(1 - \hat{p}_3)/\bar{N}_3]} = \frac{0.00747}{0.00294} = 2.536$$

Group 4: $\hat{p}_{41} = 0.16, \hat{p}_{42} = 0.20, \hat{p}_{43} = 0.16, \hat{p}_{44} = 0.00, \hat{p}_4 = 0.13,$
 $\bar{N}_4 = 25, \hat{\text{Var}}(\hat{p}_{4j}) = 0.00787, \hat{p}_4(1 - \hat{p}_4)/\bar{N}_4 = 0.00452$

$$\hat{K}_4 = \frac{\hat{\text{Var}}(\hat{p}_{4j})}{[\hat{p}_4(1 - \hat{p}_4)/\bar{N}_4]} = \frac{0.00787}{0.00452} = 1.739$$

Group 5: $\hat{p}_{51} = 0.92, \hat{p}_{52} = 0.84, \hat{p}_{53} = 0.64, \hat{p}_{54} = 0.76, \hat{p}_5 = 0.79$
 $\bar{N}_5 = 25, \hat{\text{Var}}(\hat{p}_{5j}) = 0.01427, \hat{p}_5(1 - \hat{p}_5)/\bar{N}_5 = 0.00664$

$$\hat{K}_5 = \frac{\hat{\text{Var}}(\hat{p}_{5j})}{[\hat{p}_5(1 - \hat{p}_5)/\bar{N}_5]} = \frac{0.01427}{0.00664} = 2.150$$

Group 6: $\hat{p}_{61} = \hat{p}_{62} = \hat{p}_{63} = \hat{p}_{64} = 1.0$, $\hat{p}_6 = 1.0$, $\hat{N}_6 = 25$, $\text{Var}(\hat{p}_{6j}) = 0.0$,

$$\hat{p}_6(1 - \hat{p}_6)/\bar{N}_6 = 0.0$$

\hat{K}_6 is indeterminate and so we take it to be 1.0.

Thus,

$$\hat{\bar{K}} = \frac{\sum_{i=1}^6 \hat{K}_i}{6} = \frac{0.236 + 0.362 + 2.536 + 1.739 + 2.150 + 1.000}{6} = 1.337$$

$\hat{\bar{K}}$ is the average adjustment factor which is used to adjust the observed sample sizes to effective sample sizes. The sample sizes are adjusted downward so as to maintain the observed response rates within each tank. The results of the adjustment procedure are presented in Table IX.1. These adjusted values are used as basic input "data" for subsequent analyses. We then proceed as if there is no tank to tank variation within groups. The extrabinomial variation has been accounted for by the adjustment procedure.

TABLE IX.1 EFFECTIVE SAMPLE SIZES AND RESPONSES IN HOLCOMBE AND PHIPPS
COMPOUND D FRY MORTALITY DATA AFTER ADJUSTMENT
FOR TANK TO TANK HETEROGENEITY

Group		Tank A	Tank B	Tank C	Tank D
1	Dead	1.50	1.50	0.75	0.75
	Live	17.20	17.20	17.95	17.95
	Total	18.70	18.70	18.70	18.70
2	Dead	2.24	0.75	1.50	1.50
	Live	16.45	17.95	17.20	17.20
	Total	18.70	18.70	18.70	18.70
3	Dead	1.50	0.00	3.74	0.75
	Live	17.20	18.70	14.96	17.96
	Total	18.70	18.70	18.70	18.70
4	Dead	2.99	3.74	2.99	0.00
	Live	15.70	17.96	15.70	18.70
	Total	18.70	18.70	18.70	18.70
5	Dead	17.20	15.70	11.97	14.21
	Live	1.50	2.99	6.73	4.49
	Total	18.70	18.70	18.70	18.70
6	Dead	18.70	18.70	18.70	18.70
	Live	0.00	0.00	0.00	0.00
	Total	18.70	18.70	18.70	18.70

We next illustrate the adjustment procedure on the embryo mortality data of Jarvinen for compound B. The preliminary test of tank to tank heterogeneity within treatment groups is highly significant ($Z = 2.54$, corresponding to an observed significance level of 0.005). This statistical significance is due to group 1 (control) which shows strong tank to tank differences, group 6 which shows a moderate tank to tank difference, and group 3 which shows possible indications -- but at best weak statistical evidence -- of tank to tank heterogeneity.

In this example $J = 2$, N_{ij} are close to \bar{N}_i (within 1) except for group 2. We will assume here that $N_{ij} = \bar{N}_i$ when calculating K_i 's. This assumption can be refined somewhat, if necessary, to calculate separate adjustment factors for each tank, but we will not do that here. This will await the development of adjustment procedures based on maximum likelihood estimation.

Group 1: $\hat{p}_{11} = .235, \hat{p}_{12} = .04, \hat{p}_1 = 14/101 = 0.138, N_{11} = 51, N_{12} = 50, \bar{N}_1 = 50.5$

$$\hat{\text{Var}}(\hat{p}_{ij}) = 0.019$$

$$\hat{p}_1(1 - \hat{p}_1)/\bar{N}_1 = 0.00236$$

$$\hat{K}_1 = \frac{\hat{\text{Var}}(\hat{p}_{ij})}{[\hat{p}_1(1 - \hat{p}_1)/\bar{N}_1]} = \frac{.019}{.00236} = 8.056$$

Group 2: $\hat{p}_{21} = .105, \hat{p}_{22} = .038, \hat{p}_2 = 8/109 = .073, N_{21} = 57, N_{22} = 52, \bar{N}_2 = 54.5$

$$\hat{\text{Var}}(\hat{p}_{2j}) = .0022$$

$$\hat{p}_2(1 - \hat{p}_2)/\bar{N}_2 = .00124$$

$$\hat{K}_2 = \frac{\hat{\text{Var}}(\hat{p}_{2j})}{[\hat{p}_2(1 - \hat{p}_2)/\bar{N}_2]} = \frac{.0022}{.00124} = 1.799$$

Group 3: $\hat{p}_{31} = .04, \hat{p}_{32} = 1.4, \hat{p}_3 = 9/100 = .09, N_{31} = N_{32} = \bar{N}_3 = 50$

$$\hat{\text{Var}}(\hat{p}_{3j}) = .005 \qquad \hat{p}_3(1 - \hat{p}_3)/\bar{N}_3 = .0016$$

$$\hat{K}_3 = \frac{\hat{\text{Var}}(\hat{p}_{3j})}{[\hat{p}_3(1 - \hat{p}_3)/\bar{N}_3]} = \frac{.005}{.0016} = 3.053$$

Group 4: $\hat{p}_{41} = .02, \hat{p}_{42} = .021, \hat{p}_4 = 2/98 = .020, N_{41} = 50, N_{42} = 48, \bar{N}_4 = 49$

$$\hat{\text{Var}}(\hat{p}_{4j}) = 3.4 \times 10^{-7} \qquad \hat{p}_4(1 - \hat{p}_4)/\bar{N}_4 = .00041$$

$$\hat{K}_4 = \frac{\hat{\text{Var}}(\hat{p}_{4j})}{[\hat{p}_4(1 - \hat{p}_4)/\bar{N}_4]} = \frac{3.4 \times 10^{-7}}{.00041} = 8 \times 10^{-4} = 0$$

Group 5: $\hat{p}_{51} = .077, \hat{p}_{52} = .038, \hat{p}_5 = 6/105 = .057, N_{51} = 52, N_{52} = 53, \bar{N}_5 = 52.5$

$$\hat{\text{Var}}(\hat{p}_{5j}) = .00077 \qquad \hat{p}_5(1 - \hat{p}_5)/\bar{N}_5 = .00102$$

$$\hat{K}_5 = \frac{\hat{\text{Var}}(\hat{p}_{5j})}{[\hat{p}_5(1 - \hat{p}_5)/\bar{N}_5]} = \frac{.00077}{.00102} = .750$$

Group 6: $\hat{p}_{61} = .02, \hat{p}_{62} = .137, \hat{p}_6 = 8/101 = .079, N_{61} = 50, N_{62} = 51, \bar{N}_6 = 50.5$

$$\hat{\text{Var}}(\hat{p}_{6j}) = .00687 \qquad \hat{p}_6(1 - \hat{p}_6)/\bar{N}_6 = .00144$$

$$\hat{K}_6 = \frac{\hat{\text{Var}}(\hat{p}_{6j})}{[\hat{p}_6(1 - \hat{p}_6)/\bar{N}_6]} = \frac{.00687}{.00144} = 4.77$$

Assuming that the relatively high embryo mortality response in group 1 tank 1 is not an outlier and that the inflation in variability is constant across groups, we calculate an average adjustment factor across treatment groups.

Thus,

$$\hat{\bar{K}} = \frac{\sum_{i=1}^6 K_i}{6} = \frac{8.056 + 1.799 + 3.053 + 0 + 0.750 + 4.77}{6} = 3.071$$

$\hat{\bar{K}}$ is the average adjustment factor across groups. Alternatively we might use the separate adjustment factors within groups. The results of the adjustment procedure are presented in Table IX.2. These adjusted values can be used as basic input "data" for subsequent analyses. We proceed with further analyses as if there is no tank to tank heterogeneity within groups.

TABLE IX.2 EFFECTIVE SAMPLE SIZES AND RESPONSES IN
JARVINEN COMPOUND B EMBRYO MORTALITY
DATA AFTER ADJUSTMENT FOR TANK
TO TANK HETEROGENEITY

Group		Tank A	Tank B
1	Dead	3.91	0.65
	Live	12.70	15.63
	Total	16.61	16.28
2	Dead	1.95	0.65
	Live	16.61	16.28
	Total	18.56	16.93
3	Dead	0.65	2.28
	Live	15.63	14.00
	Total	16.28	16.28
4	Dead	0.33	0.33
	Live	15.95	15.30
	Total	16.28	15.63
5	Dead	1.30	0.65
	Live	15.63	16.61
	Total	16.93	17.26
6	Dead	0.33	2.28
	Live	15.95	14.33
	Total	16.28	16.61

X. OUTLIER DETECTION PROCEDURES

A. Background

Another preliminary analysis of importance is the detection of responses which do not appear to be in conformance with the substantial majority of responses. Such exceptional responses are often referred to as "outliers". Outlier detection procedures are used to decide how extreme a response must be in order to rule out the possibility that its value is reasonably likely to be due just to random variation. Consider for example the percentage embryo mortality responses from DeFoe's test on compound C that are displayed in Figure VI.1. We remarked that the mortality rate in group 2, tank A appears to be widely separated from the others. Can such a separation be explained by random variation or is there some systematic factor peculiar to this tank? Similarly, the percentage embryo mortality observed in group 1, tank A in Jarvinen's test on compound B is widely separated from the other responses. Can a separation of this magnitude be reasonably explained by random variation or is there some systematic factor peculiar to this tank?

Barnett and Lewis [23] describe a wide class of outlier detection procedures, to screen out those extreme responses that cannot be reasonably attributed to random variation. They include a procedure appropriate for binomial responses (section 3.4, pp 122-124). Their procedure is based on the assumption of n independent responses X_1, \dots, X_n , each binomially distributed with parameters m and p . They base their outlier test on the exact conditional distribution of $\max X_j$ given $\sum_j X_j$. In our data n represents the number of tanks per group, m is the number of embryos or fry per tank (assumed to be equal from tank to tank), and X_j is the number of responses (e.g. dead embryos) per tank. Their tabulation, Table XIX (pp 320-322) includes only the range of values $n > 3$, $m \geq 10$, $X_{(n)} = m, m-1, m-2$. This is quite inadequate for the ranges of parameters and responses that arise in toxicity tests. Thus their exact conditional test is not too useful for our needs.

Barnett and Lewis state, on page 123, that an alternative, approximate approach to outlier detection in the binomial case is to transform $\{X_j/m\}$ using the arc sine transformation and then apply normal theory based procedures to these transformed values. This approach, and variants on it, are in the spirit of the methods that we recommend in the remainder of this section. We consider both graphical and numerical procedures.

The theoretical bases of our suggested methods are discussed in Appendix AX.

B. Application of Outlier Detection Procedures to Fish Toxicity Data

We apply the transformations discussed in Appendix AX to construct graphical outlier detection procedures based on normal probability plotting techniques and associated formal outlier detection tests. We apply these procedures to the following situations:

DeFoe: compound C
embryo mortality data
fry mortality data

Holcombe and Phipps: Compound D
embryo mortality data
fry mortality data

Jarvinen: compound B
embryo mortality data

DeFoe compound C

I = 6, J = 2

(i.e. 6 groups, 2 tanks per group).

Embryo Mortality

Group 1 $\hat{p} = .41$; all expected frequencies are greater than 5.
 Group 2 $\hat{p} = .51$; all expected frequencies are greater than 5.
 Group 3 $\hat{p} = .37$; all expected frequencies are greater than 5.
 Group 4 $\hat{p} = .37$; all expected frequencies are greater than 5.
 Group 5 $\hat{p} = .39$; all expected frequencies are greater than 5.
 Group 6 $\hat{p} = .38$; all expected frequencies are greater than 5.

We apply the transformation suggested in case 1. Data summaries are given below.

Calculate $\left(1 - \frac{N_j}{N}\right)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}}\right)$ for each tank within each group.

	X_j	N_j	\hat{p}	$N_j \hat{p}$	$(N_j \hat{p} \hat{q})^{1/2}$	N_j/N	$\left(1 - \frac{N_j}{N}\right)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}}\right)$
<u>Group 1</u>							
Tank A	22	50	0.41	20.5	3.478	.50	.610
Tank B	19	50	0.41	20.5	3.478	.50	-.610
<u>Group 2</u>							
Tank A	32	50	0.51	25.5	3.535	.50	2.600
Tank B	19	50	0.51	25.5	3.535	.50	-2.600
<u>Group 3</u>							
Tank A	20	50	0.37	18.5	3.414	.50	0.621
Tank B	17	50	0.37	18.5	3.414	.50	-0.621
<u>Group 4</u>							
Tank A	16	50	0.37	18.5	3.414	.50	-1.036
Tank B	21	50	0.37	18.5	3.414	.50	1.036
<u>Group 5</u>							
Tank A	22	50	0.39	19.5	3.449	.50	1.025
Tank B	17	50	0.39	19.5	3.449	.50	-1.025
<u>Group 6</u>							
Tank A	19	50	0.38	19.0	3.432	.50	0
Tank B	19	50	0.38	19.0	3.432	.50	0

To prepare the normal probability plot we order the standardized values and plot the i -th smallest against the plotting position $100 \times (i - 0.5)/12$ on the probability scale. These values are indicated below.

i	1	2	3	4	5	6
Ordered Value	-2.600	-1.036	-1.025	-0.621	-0.610	0
Plotting Position	4.2	12.5	20.8	29.2	37.5	45.8

i	7	8	9	10	11	12
Ordered Value	0	0.610	0.621	1.025	1.036	2.600
Plotting Position	54.2	62.5	70.8	79.2	87.5	95.8

The normal probability plot of these points is shown in Figure X.1. The plot appears perfectly symmetrical about 0 since $J = 2$ and so the responses within groups have correlation -1.0 . The effective sample size is thus $6(J - 1) = 6(2 - 1) = 6$ independent observations. The reference line in the plot is the standard normal distribution function. If the response rates are homogeneous within groups then the standardized values should lie near this line. If there is extrabinomial variation in the data, that is random tank to tank variation within groups, then the points should lie along a line or a curve with steeper slope than the standard distribution function. If there are outlying responses in the data then they should be far removed from the line or curve that typifies the bulk of the data. This latter situation is seen to be the case. The bulk of the data lie very nicely along the standard normal c.d.f. line. The values corresponding to group 2 are far removed from this line.

To determine the extent of statistical evidence that the apparent outliers did not occur just due to chance we calculate the probability that the maximum absolute value of six independent standard normal random variables exceeds 2.600. More precisely let Z_1, Z_2, \dots, Z_6 be six independent standard normal random variables. Then

$$\begin{aligned}
 P[\max_{j=1, \dots, 6} |Z_j| \geq 2.600] &= P[\text{at least one } |Z_j| \geq 2.600] \\
 &= 1 - P[\text{all } |Z_j| \leq 2.600] = 1 - \{P[|Z_1| \leq 2.600]\}^6 = 1 - (.9907)^6 \\
 &= 0.055
 \end{aligned}$$

This is of borderline statistical significance. We can thus infer that based on this test there is marginal statistical evidence that group 2 contains an outlying tank.

The appearance of Figure VI.1 suggests that the response from Tank 2A is more than a marginal outlier. We can increase the sensitivity of the above outlier test by incorporating additional information. If we assume that there is no trend in response rates across groups then we can estimate the response rate based on all 12 tanks and can ignore the correction factor $(1 - N_j/N)^{-1/2}$. In general this assumption will not hold but it seems reasonable in this example based on the appearance of Figure VI.1 and on toxicological considerations (i.e. relatively little penetration of chemical into the embryo). The value of \hat{p} based on 12 tanks is $243/600 = 0.405$. The largest standardized value is that from Tank 2A, namely

$$Z = \frac{32 - (50)(.405)}{[50(.405)(.595)]^{1/2}} = 3.385$$

What is the probability that the most extreme of 11 independent standard normal random variable exceeds 3.385 in absolute value?

$$P[\max_{j=1, \dots, 12} |Z_j| \geq 3.385] = 1 - \{P[|Z_1| \leq 3.385]\}^{11} = 1 - (.9997)^{11} \\ = 0.003$$

We can thus infer that, with the additional assumption of no trend in response with increasing treatment level, there is strong statistical evidence that the response rate in Tank 2A is an outlying value.

Note that just because the Tank 2A response is an outlier does not in and of itself mean that the data should be discarded or disregarded. Rather, the investigator needs to reexamine the records for this tank to determine the reason for the atypical response. If it is due to clerical error, to experimental mishap, to outbreak of a disease unrelated to the toxicant, etc then perhaps the Tank 2A response is inappropriate and should be disregarded. If it represents normal biological variation then the response should be considered with the others. This is a matter for biological judgement. Outlier detection procedures are merely screening devices to direct attention to those places where such biological judgement need be applied.

DeFoe compound C - Fry mortality

Group 1	$\hat{p} = 0;$
Group 2	$\hat{p} = 0;$
Group 3	$\hat{p} = 0.05;$
Group 4	$\hat{p} = 0.024;$
Group 5	$\hat{p} = 0.225;$ expected frequencies less than 5 in both tanks
Group 6	$\hat{p} = 1.00$ ($\hat{q} = 0$);

Thus groups 1, 2, 3, 4, 6 correspond to the Poisson case 2. (In group 6 we interchange the roles of p and q). Group 5 corresponds to case 3.

For all groups but 5 we calculate $\left(1 - \frac{N_j}{N}\right)^{-1/2} 2 \left[\sqrt{X_j} - \sqrt{N_j \hat{p}} \right]$. For group 5 we carry out an arc sine transformation

	X_j	N_j	\hat{p}	$N_j \hat{p}$	N_j/N	$\left(1 - N_j/N\right)^{-1/2} 2 \left[\sqrt{X_j} - \sqrt{N_j \hat{p}} \right]$
--	-------	-------	-----------	---------------	---------	--

Group 1

Tank A	0	20	0	0	.50	0
Tank B	0	20	0	0	.50	0

Group 2

Tank A	0	20	0	0	.50	0
Tank B	0	20	0	0	.50	0

Group 3

Tank A	0	20	.05	1.0	.50	-2.828
Tank B	2	20	.05	1.0	.50	1.172

Group 4

Tank A	0	21	.024	0.512	0.512	-2.049
Tank B	1	20	.024	0.488	0.488	0.843

	X_j	N_j	\hat{q}	$N_j \hat{q}$	N_j/N	$\left(1 - N_j/N\right)^{1/2} 2 \left[\sqrt{X_j} - \sqrt{N_j \hat{q}} \right]$
--	-------	-------	-----------	---------------	---------	---

Group 6

Tank A	0	20	0	0	.50	0
Tank B	0	20	0	0	.50	0

	X_j	N_j	\hat{p}	$N_j \hat{p}$	N_j/N	\hat{p}_j	$\left(1 - N_j/N\right)^{-1/2} 2 \sqrt{N_j} \left[\arcsin \sqrt{\hat{p}_j} - \arcsin \sqrt{\hat{p}} \right]$
--	-------	-------	-----------	---------------	---------	-------------	---

Group 5

Tank A	4	20	0.225	4.5	0.50	0.20	-0.387
Tank B	5	20	0.225	4.5	0.50	0.25	0.372

To prepare the normal probability plot we order the standardized values and plot the i -th smallest against the plotting position $100 \times (i - 0.5)/12$ on the probability scale. These values are indicated below.

i	1	2	3	4	5	6	7	8	9
Ordered Value	-2.828	-2.049	-0.387	0	0	0	0	0	0
Plotting Position	4.2	12.5	20.8	29.2	37.5	45.8	54.2	62.5	70.8

i	10	11	12
Ordered Value	0.372	0.843	1.172
Plotting Position	79.2	87.5	95.8

The normal probability plot of these points appears in Figure X.2. We note two points well below the standard $N(0, 1)$ line. These points correspond to Tanks 3A, 4A. Both correspond to frequencies of 0, where the normal approximation is least reasonable. Furthermore, their companion tanks do not show up as outliers. Thus before we say there is an outlier, we should compare the proportions in the two replicate tanks to see if there is any statistical evidence of differences.

There is an exact test for the equality of two Poisson means. Nelson [24] discusses this test in detail.

In order to use this test for detecting outlier tanks we need to test a slightly more general hypothesis. Namely consider the 2×2 table.

		Replicate	
		A	B
Live	X	Y	
Dead			
	N_A	N_B	N

If $\frac{X}{N_A} < .1$, $\frac{Y}{N_B} < .1$ or if $\frac{X}{N_A} > .9$, $\frac{Y}{N_B} > .9$ then we're in the Poisson case.

Now $\lambda_A = N_A P_A$, $\lambda_B = N_B P_B$ implies that if $P_A = P_B$ then $\lambda_A / \lambda_B = N_A / N_B$.

We thus wish to test the hypothesis

$$H_0: \frac{\lambda_A}{\lambda_B} = \rho \quad \text{where } \rho = \frac{N_A}{N_B}$$

Nelson's test rejects H_0 at level α if

$$\frac{Y}{X+1} \geq \frac{1}{\rho} F(2X+2, 2Y; 1-\alpha/2)$$

or if

$$\frac{Y+1}{X} \leq \frac{1}{\rho} F(2X, 2Y+2; \alpha/2) \equiv \frac{1}{\rho} \frac{1}{F(2Y+2, 2X; 1-\alpha/2)}$$

where $F(v_1, v_2; \gamma)$ represents the upper γ point of the F c.d.f. with v_1, v_2 degrees of freedom. If $Y = 0, X > 0$ we can only carry out the one sided test. (We have just one sided information concerning λ_B).

$$H_0: \lambda_A / \lambda_B = \rho \quad \text{vs} \quad H_1: \lambda_A / \lambda_B > \rho$$

Nelson's test rejects H_0 at level α if $X \geq \rho F(2, 2X; 1-\alpha)$.

We now apply this Poisson test to the outlier detection problem. Consider groups 3, 4. These give rise to the two extreme points on the plot: -2.828, -2.049. Let us see if these should be regarded as outliers.

Replicate

In Group 3 we have

	A	B	
Dead	0	2	2
Live	20	18	38
	20	20	40

$$\text{Thus } X = 0 \quad X \sim P_0(\lambda_A) \equiv P_0(20p_A)$$

$$Y = 2 \quad Y \sim P_0(\lambda_B) \equiv P_0(20p_B)$$

Since $X = 0$, we can carry out only a one sided test. Note that $\rho = 1$.

$$H_0: \lambda_B = \lambda_A$$

$$H_1: \lambda_B > \lambda_A$$

We reject H_0 at level α if $Y > F(2, 2Y; 1-\alpha)$. In our example $Y = 2$. The critical value, $F(2, 4; .95)$ is 6.94, which exceeds 2. There

is thus no statistical evidence of differences among the two responses in group 3. Thus Tank 3 A is not an outlier.

We now carry out the test for group 4. The situation is less extreme than that in group 3, however we go through with the test for illustrative purposes.

		Replicate		
		A	B	
In Group 4 we have	Dead	0	1	1
	Live	21	19	40
		21	20	41

$$\text{Thus } X = 0 \quad X \sim P_o(\lambda_A) \equiv P_o(21p_A)$$

$$Y = 1 \quad Y \sim P_o(\lambda_B) \equiv P_o(20p_B)$$

Since $X = 0$, we can carry out only a one sided test.

$$H_o: \frac{\lambda_B}{\lambda_A} = \frac{N_B}{N_A} \equiv \frac{20}{21} = 0.952 \equiv \rho.$$

$$H_1: \frac{\lambda_B}{\lambda_A} > 0.952$$

From our previous discussion (interchanging the roles of X and Y)

$$\begin{aligned} \text{reject } H_o \text{ if } Y \geq \rho F(2, 2Y; .95) &= \frac{N_B}{N_A} F(2, 2Y; .95) \\ &= .952F(2, 2Y; .95) \end{aligned}$$

$$\text{In our case } Y = 1, \quad .952F(2, 2Y; .95) = (.952)(19.0) = 18.088$$

Thus we cannot reject H_o . There is no statistical evidence of differences in response rates among the tanks.

Thus tank 4A is not an outlier.

Holcombe and Phipps compound D

I = 6, J = 4

Embryo Mortality

(i.e. 6 groups, 4 tanks per group)

Embryo mortality

Group 1 $\hat{p} = 0.35$; all expected frequencies greater than 5.
 Group 2 $\hat{p} = 0.35$; all expected frequencies greater than 5.
 Group 3 $\hat{p} = 0.315$; all expected frequencies greater than 5.
 Group 4 $\hat{p} = 0.39$; all expected frequencies greater than 5.
 Group 5 $\hat{p} = 0.335$; all expected frequencies greater than 5.
 Group 6 $\hat{p} = 0.38$; all expected frequencies greater than 5.

We apply the transformation suggested in case 1. Data summaries are given below.

Calculate $\left(1 - \frac{N_j}{N}\right)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}}\right)$ for each tank within each group.

	X_j	N_j	\hat{p}	$N_j \hat{p}$	$\sqrt{N_j \hat{p} \hat{q}}$	N_j/N	$\left(1 - N_j/N\right)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}}\right)$
<u>Group 1</u>							
Tank A	17	50	0.35	17.5	3.373	0.25	-0.171
Tank B	21	50	0.35	17.5	3.373	0.25	1.198
Tank C	12	50	0.35	17.5	3.373	0.25	-1.883
Tank D	20	50	0.35	17.5	3.373	0.25	0.856
<u>Group 2</u>							
Tank A	19	50	0.35	17.5	3.373	0.25	0.514
Tank B	14	50	0.35	17.5	3.373	0.25	-1.198
Tank C	16	50	0.35	17.5	3.373	0.25	-0.514
Tank D	21	50	0.35	17.5	3.373	0.25	1.198
<u>Group 3</u>							
Tank A	15	50	0.315	15.75	3.293	0.25	-0.263
Tank B	12	50	0.315	15.75	3.293	0.25	-1.315
Tank C	24	50	0.315	15.75	3.293	0.25	2.893
Tank D	12	50	0.315	15.75	3.293	0.25	-1.315

	X_j	N_j	\hat{p}	$N_j \hat{p}$	$N_j \hat{p} \hat{q}$	$\sqrt{N_j/N}$	$(1 - N_j/N)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}} \right)$
<u>Group 4</u>							
Tank A	20	50	0.39	19.5	3.449	0.25	0.167
Tank B	21	50	0.39	19.5	3.449	0.25	0.502
Tank C	23	50	0.39	19.5	3.449	0.25	1.172
Tank D	14	50	0.39	19.5	3.449	0.25	-1.841
<u>Group 5</u>							
Tank A	18	50	0.335	16.75	3.337	0.25	0.433
Tank B	19	50	0.335	16.75	3.337	0.25	0.779
Tank C	14	50	0.335	16.75	3.337	0.25	-0.952
Tank D	16	50	0.335	16.75	3.337	0.25	-0.260
<u>Group 6</u>							
Tank A	14	50	0.38	19.0	3.432	0.25	-1.682
Tank B	16	50	0.38	19.0	3.432	0.25	-1.009
Tank C	25	50	0.38	19.0	3.432	0.25	2.019
Tank D	21	50	0.38	19.0	3.432	0.25	0.673

To prepare the normal probability plot we order the standardized values and plot the i -th smallest against the plotting position $100 \times (1 - 0.5)/24$ on the probability scale. These values are indicated below

i	1	2	3	4	5	6	7	8
Ordered Value	-1.883	-1.841	-1.682	-1.315	-1.315	-1.198	-1.009	-0.952
Plotting Position	2.1	6.2	10.4	14.6	18.7	22.9	27.1	31.2
i	9	10	11	12	13	14	15	16
Ordered Value	-0.514	-0.263	-0.260	-0.171	0.167	0.433	0.502	0.514
Plotting Position	35.4	39.6	43.7	47.9	52.1	56.2	60.4	64.6

AD-A134 389

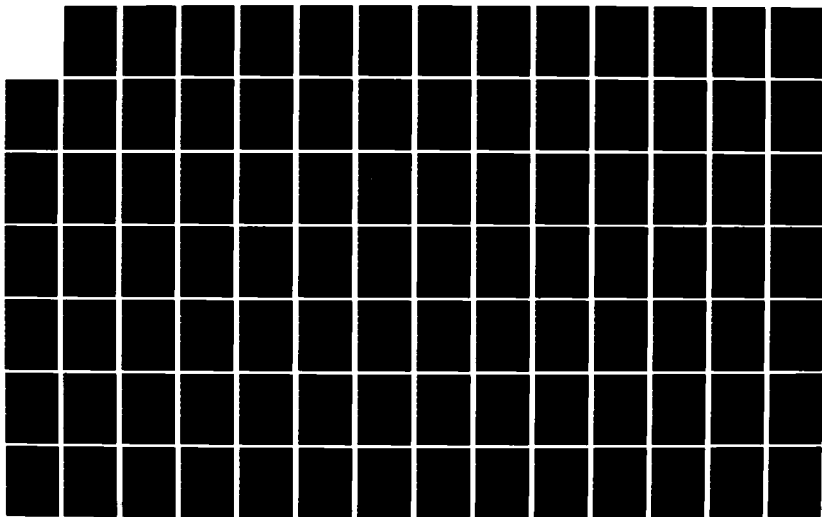
DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF
TOXICITY(U) OHIO STATE UNIV RESEARCH FOUNDATION
COLUMBUS P I FEDER ET AL. MAY 81 DAMD17-79-C-9150

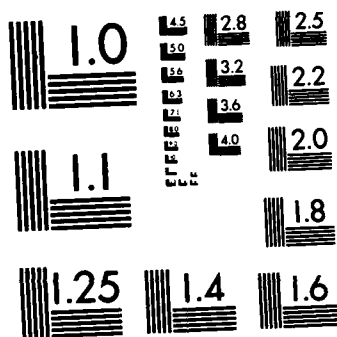
2/5

UNCLASSIFIED

F/G 6/20

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

i	17	18	19	20	21	22	23	24
Ordered Value	0.673	0.779	0.856	1.172	1.198	1.198	2.019	2.893
Plotting Position	68.7	72.9	77.1	81.2	85.4	89.6	93.7	97.9

The normal probability plot of these points is shown in Figure X.3. Due to the within group correlation, the effective sample size is $6(J - 1) = 6(4 - 1) = 18$ independent observations. Is the Group 3, Tank C response an outlier? To determine the extent of statistical evidence that the apparent outlier did not occur just due to chance we calculate the probability that the maximum absolute value of 18 independent standard normal random variables does not exceed 2.893. Let Z_1, Z_2, \dots, Z_{18} be 18 independent standard normal random variables.

$$P[\max_{j=1, \dots, 18} |Z_j| \geq 2.893] = 1 - P[\text{all } |Z_j| \leq 2.893] = 1 - (.9962)^{18} \\ = 0.066.$$

This is of borderline statistical significance. We can thus infer that there is marginal statistical evidence that Group 3 contain an outlying tank.

Note that from the scatter plot of embryo mortality vs treatment in Figure VI.5. It is clear that there is no trend in the data and that the Tank 3C response does not stand out from those of the six groups as a whole. If we knew that there was no trend over groups we could construct a more powerful test by pooling all the tanks and calculating a common \hat{p} . However there is no point in doing this since we see from the scatter plot that Tank 3C is not out of line with respect to the pooled responses, but rather just with those in Group 3. The reason for this, if any, might be pursued.

Holcombe and Phipps compound D

$$I = 6, J = 4$$

(i.e. 6 groups, 4 tanks per group)

Fry Mortality

$$\text{Group 1 } \hat{p} = 0.06$$

$$\text{Group 2 } \hat{p} = 0.08$$

$$\text{Group 3 } \hat{p} = 0.08$$

$$\text{Group 4 } \hat{p} = 0.13; \quad \text{expected frequencies less than 5 in each tank.}$$

$$\text{Group 5 } \hat{p} = 0.79; \quad \text{all expected frequencies greater than 5.}$$

$$\text{Group 6 } \hat{p} = 1.00 (\hat{q} = 0).$$

Thus groups 1, 2, 3, 6 correspond to Poisson case 2. Group 4 corresponds to case 3 (possibly to case 2). Group 5 corresponds to case 1.

Thus for groups 1, 2, 3,
we calculate

$$(1 - N_j/N)^{-1/2} 2 [\sqrt{X_j} - \sqrt{N_j \hat{p}}].$$

For group 6 we calculate

$$(1 - N_j/N)^{-1/2} 2 [\sqrt{N_j - X_j} - \sqrt{N_j \hat{q}}].$$

For group 4 we calculate

$$(1 - N_j/N)^{-1/2} 2\sqrt{N_j} [\arcsin\sqrt{\hat{p}_j} - \arcsin\sqrt{\hat{p}}]$$

For group 5 we calculate

$$(1 - N_j/N)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}} \right)$$

	X_j	N_j	\hat{p}	$N_j \hat{p}$	N_j/N	$(1 - N_j/N)^{-1/2}$	$2 [\sqrt{X_j} - \sqrt{N_j \hat{p}}]$
<u>Group 1</u>							
Tank A	2	25	.06	1.5	.25		0.438
Tank B	2	25	.06	1.5	.25		0.438
Tank C	1	25	.06	1.5	.25		-0.519
Tank D	1	25	.06	1.5	.25		-0.519
<u>Group 2</u>							
Tank A	3	25	.08	2.0	.25		0.734
Tank B	1	25	.08	2.0	.25		-0.957
Tank C	2	25	.08	2.0	.25		0
Tank D	2	25	.08	2.0	.25		0
<u>Group 3</u>							
Tank A	2	25	.08	2.0	.25		0
Tank B	0	25	.08	2.0	.25		-3.266
Tank C	5	25	.08	2.0	.25		1.898
Tank D	1	25	.08	2.0	.25		-0.957

$$(1 - N_j/N)^{-1/2} 2\sqrt{N_j} \left[\arcsin \sqrt{\hat{p}_j} - \arcsin \sqrt{\hat{p}} \right]$$

$$X_j \quad N_j \quad \hat{p} \quad N_j \hat{p} \quad N_j/N \quad \hat{p}_j \quad \arcsin \sqrt{\hat{p}}$$

Group 4

Tank A	4	25	.13	3.25	.25	.16	.493
Tank B	5	25	.13	3.25	.25	.20	1.094
Tank C	4	25	.13	3.25	.25	.16	.493
Tank D	0	25	.13	3.25	.25	0	-4.259

$$X_j \quad N_j \quad \hat{p} \quad N_j \hat{p} \quad N_j/N \quad \sqrt{N_j \hat{p} \hat{q}} \quad (1 - N_j/N)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{\sqrt{N_j \hat{p} \hat{q}}} \right)$$

Group 5

Tank A	23	25	.79	19.75	.25	2.037	1.842
Tank B	21	25	.79	19.75	.25	2.037	0.709
Tank C	16	25	.79	19.75	.25	2.037	-2.126
Tank D	19	25	.79	19.75	.25	2.037	-.425

$$N_j - X_j \quad N_j \quad \hat{q} \quad N_j \hat{q} \quad N_j/N \quad (1 - N_j/N)^{-1/2} 2 \left[\sqrt{N_j - X_j} - \sqrt{N_j \hat{q}} \right]$$

Group 6

Tank A	0	25	0	0	.25	0
Tank B	0	25	0	0	.25	0
Tank C	0	25	0	0	.25	0
Tank D	0	25	0	0	.25	0

To prepare the normal probability plot we order the standardized values and plot the i -th smallest against the plotting position $100 \times (i - 0.5)/24$ on the probability scale. These values are indicated below.

	1	2	3	4	5	6	7
Ordered Value	-4.259	-3.266	-2.126	-0.957	-0.957	-0.519	-0.519
Plotting Position	2.1	6.2	10.4	14.6	18.7	22.9	27.1

	8	9	10	11	12	13	14	15	16
Ordered Value	-0.425	0	0	0	0	0	0	0	0.438
Plotting Position	31.2	35.4	39.6	43.7	47.9	52.1	56.2	60.4	64.6

	17	18	19	20	21	22	23	24
Ordered Value	0.438	0.493	0.493	0.709	0.734	1.094	1.842	1.898
Plotting Position	68.7	72.9	77.1	81.2	85.4	89.6	93.7	97.9

The normal probability plot appears in Figure X.4. It appears that the lowest 3 points are well below the $N(0, 1)$ line. These three points correspond to Tanks 4D, 3B, 5C. Tanks 4D, 3B each have observed frequencies of 0. This is the region where the normal approximation is the poorest. Thus before we say that there are any outliers, we should compare the tanks within the treatment groups using a more appropriate exact test.

First let's look at tank 3B.

Group 3:

	Replicate				
	A	B	C	D	
Dead	2	0	5	1	8
Live	23	25	20	24	92
	25	25	25	25	100

Tank 3 B is the suspected outlier. Let's compare its results to those in the other 3 tanks.

	Replicate		
	B	A, C, D	
Dead	0	8	8
Live	25	67	92
	25	75	100

We can carry out an exact test by means of the Fisher - Irwin test. (See Lehmann [25], section 4.5, Lieberman and Owen [26]).

To carry out the Fisher - Irwin test we adopt the following notational identifications for the table.

	Grp 1	Grp 2	
Spec.	X		k
Ordin.			K - k
	n	N - n	N

Where

$$k < N - k, n < N - n, k < n.$$

i.e. \bar{k} is the smallest marginal entry

\bar{n} is the smallest marginal entry in the other margin.

X is the cell entry in the cell corresponding to the (n, k) marginal categories.

In our example

	B	A, C, D	
Dead	x = 0	8	8 = k ("special")
Live	25	67	92
	25	75	100 = N
	"		
	n		

Thus $k = 8$, $n = 25$, $N = 100$, $X = x = 0$. We enter the Lieberman and Owen hypergeometric distribution tables at these parameters. We obtain $P(X < 0) = 0.091$. Thus a two sided probability is 0.182. This is quite marginal, at most.

Now group B was not chosen a priori, but rather as the most extreme of the 4 responses. Thus to get a feeling for how extreme this behavior is we carry out the following approximate calculation. $P(\text{most extreme of 4 independent responses is more significant than 0.182 level}) = 1 - P(\text{all 4 responses less significant than 0.182 level}) = 1 - (1 - 0.182)^4 = 0.55$. We thus conclude that there is no statistical evidence that the response rate in Tank 3B differs significantly from the responses rates in the other tanks in that group. We conclude that the extreme behavior of the standardized value is due to the inapplicability of the normalizing square root transformation when $X_j = 0$.

We now consider the responses in Group 4.

Group 4

	Replicate				
	A	B	C	D	
Dead	4	5	4	0	13
Live	21	20	21	25	87
	25	25	25	25	100

Tank 4D is the suspected outlier. Let's compare its results to those in the other 3 tanks.

	D	A, B, C	
	0 = x	13	13 = k
Dead	0 = x	13	13 = k
Live	25	62	87
	25	75	100 = N
	n		

Thus here $k = 13$, $n = 25$, $N = 100$, $X = x = 0$.

Entering the Lieberman and Owen tables we find that

$P(X \leq 0) = 0.018$ this is a one sided probability

Thus the observed two sided significance level = $2(0.018) = 0.036$

Now tank D was not chosen a priori. Taking selection into account we have

P (most extreme of 4 tanks more significant than 0.036 level) =

$1 - P$ (all 4 tanks less significant than 0.036 level)

= $1 - (1 - .036)^4 = 0.14$.

There is thus at most a marginal suggestion that the response rate in Tank 4D differs significantly from the response rates in the other tanks in that group. The very extreme appearance of the standardized value on the normal probability plot is again due to the inapplicability of the normalizing transformation when $X_j = 0$.

We now consider the responses in Group 5.

Group 5

	Replicate				
	A	B	C	D	
Dead	23	21	16	19	79
Live	2	4	9	6	21
	25	25	25	25	100

Tank 5C is the suspected outlier. Let's compare its results to those in the other three tanks.

	C	A, B, D	
Dead	16	63	79
Live	9 = x	12	21 = k
	25	75	100 = N
	"		
	n		

$n\hat{p} = 5.25$

In this case the "special" category is "live". Thus $k = 21$, $n = 25$, $N = 100$, $X = x = 9$. Entering the Lieberman and Owen tables we find that

$$P(X \geq 9) = 1 - P(X \leq 8) = 1 - .9638 = .0362.$$

Thus the two sided significance probability is $2(.0362) = .0724$. This is at best marginal. Now tank D was not chosen a priori. Taking selection into account, we have

$$P(\text{most extreme of 4 tanks most significant than } .0724 \text{ level}) = 1 - (1 - .0724)^4 = .260$$

Thus tank D is not significantly different than the others. Since the expected frequencies are fairly large in this example we can also carry out an asymptotic test.

	C	A, B, D	
Dead	16 (19.65)	63 (59.25)	79
Live	9 (5.25)	12 (15.75)	21
	25	75	100

Expected frequencies are in parentheses.

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(16 - 19.65)^2}{19.65} + \frac{(63 - 59.25)^2}{59.25} + \frac{(9 - 5.25)^2}{5.25} + \frac{(12 - 15.75)^2}{15.75}$$

$$\text{Thus } \chi^2 = 4.53$$

Under the hypothesis of homogeneity, χ^2 is distributed as χ_1^2 .

Thus χ^2 is significant at the 0.033 level, in close agreement with the results based on the Fisher - Irwin test, namely lack of statistical evidence of differences when selection is accounted for.

Since the normalizing transformation is appropriate for the range of responses in Group 5, we can also regard the normalized value from Tank C, -2.126, as the minimum of $6 \times (4 - 1) - 2 = 16$ standard normal deviates (we disregard those responses corresponding to Tanks 3B, 4D). The probability that a standard normal deviate is less than -2.126 is 0.017. The probability that the minimum of 22 independent standard normal deviates is less than -2.126 is thus $1 - (1 - 0.017)^{16} = 0.25$. Again there is no statistical evidence that this value is an outlier.

Jarvinen compound B

I = 6, J = 2 (i.e. 6 groups, 2 tank per group)

Embryo Mortality

The normal probability plot of standardized values (based on the case 1 transformation) is shown in Figure X.5. The plot appears perfectly symmetric about 0 since J = 2. The effective sample size is $6(J - 1) = 6$ independent observations. The reference line in the plot is the standard normal distribution function.

The behavior of the plot suggests the presence of extrabinomial variation (i.e. random tank to tank variation within groups) in the data rather than outliers. This is seen by the fact that the points lie along a curve with steeper slope than that of the standard normal distribution

function. The extreme points are not outliers since they lie directly on the curve determined by the other values. Thus the conjecture made in section VIII concerning the presence of outliers is not borne out. Note that this behavior is directly opposite to that observed in Figure X.1.

If we draw a line through the plotted points in Figure X.5 the estimated standard deviation (corresponding to the difference between the 84th and 50th percentiles) is about 1.7. Thus the estimated variance is $(1.7)^2 = 2.9$. This value is very close to the factor $K = 3.07$ that we utilized in section IX to adjust these data.

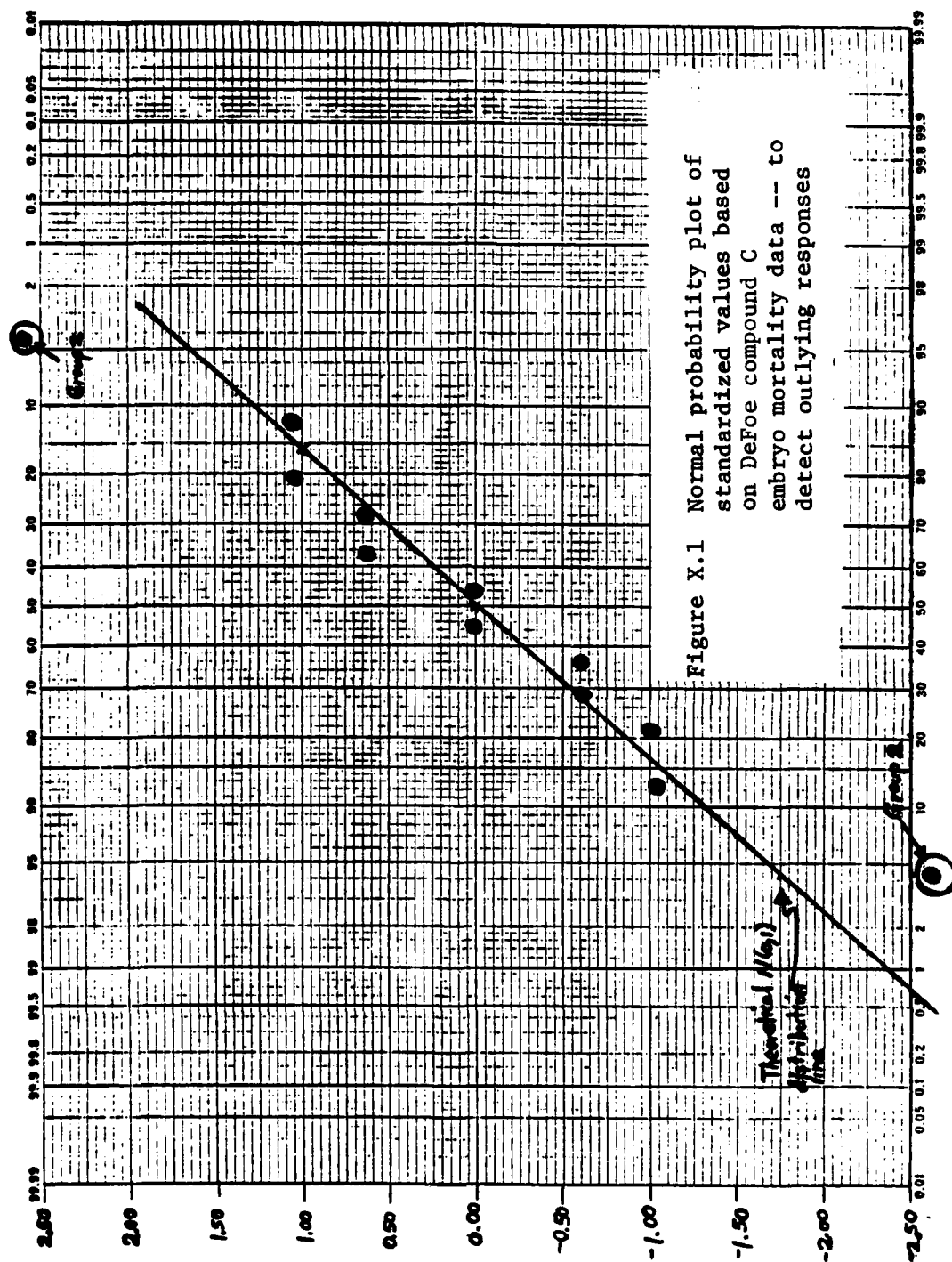


Figure X.1 Normal probability plot of standardized values based on DeFoe compound C embryo mortality data -- to detect outlying responses

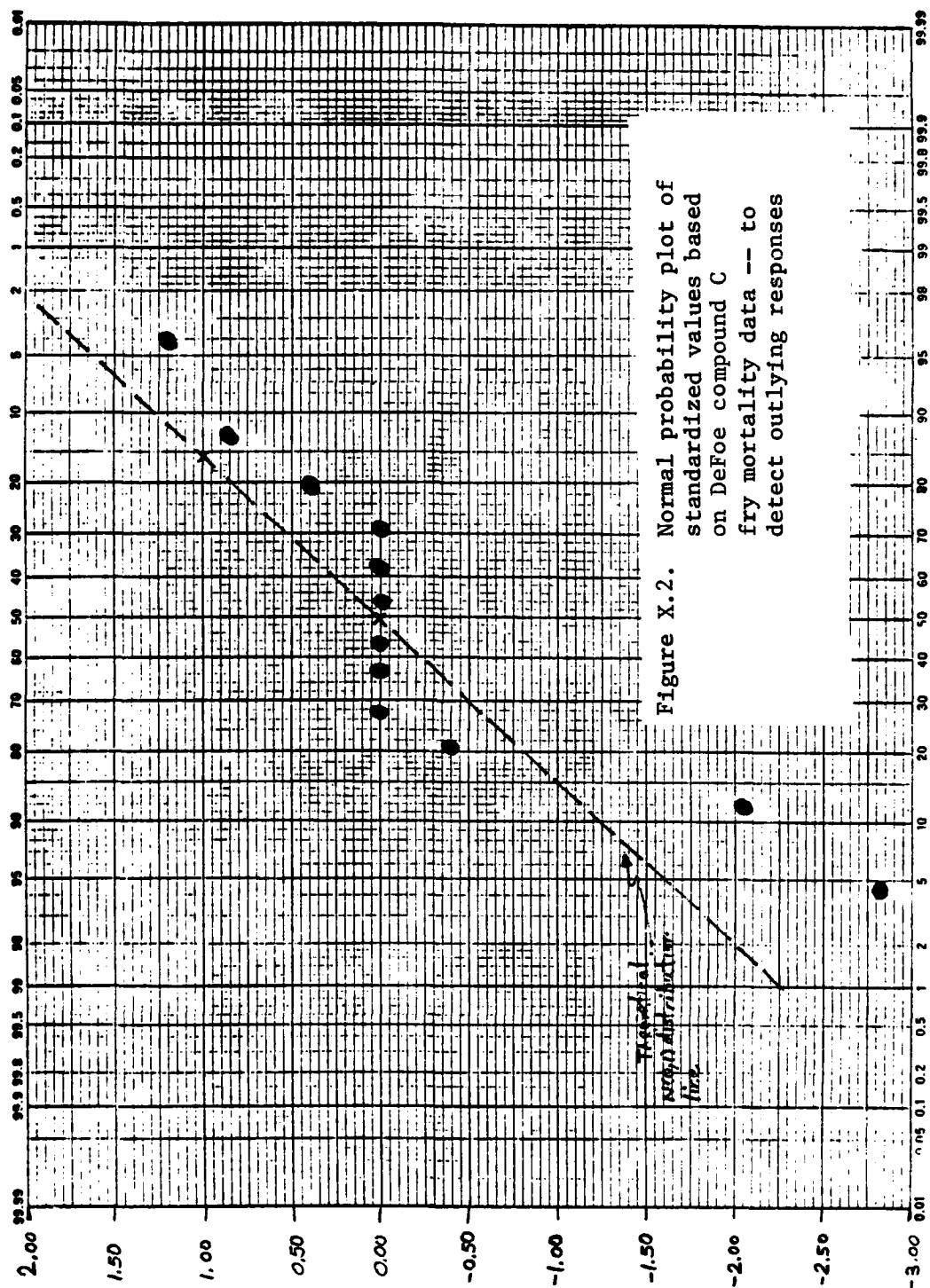


Figure X.2. Normal probability plot of standardized values based on DeFoe compound C fry mortality data -- to detect outlying responses

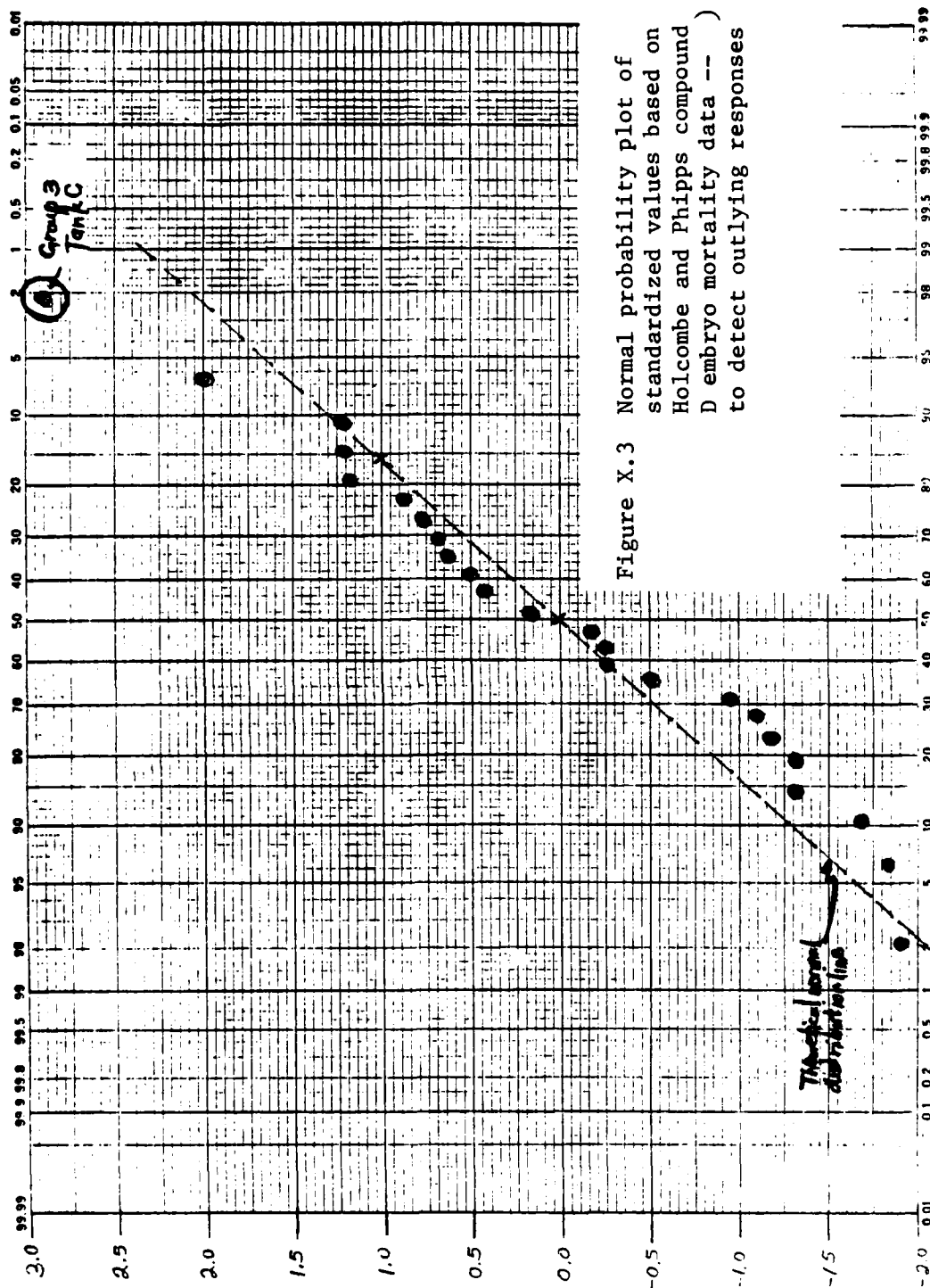


Figure X.3 Normal probability plot of standardized values based on Holcombe and Phipps compound D embryo mortality data (---) to detect outlying responses

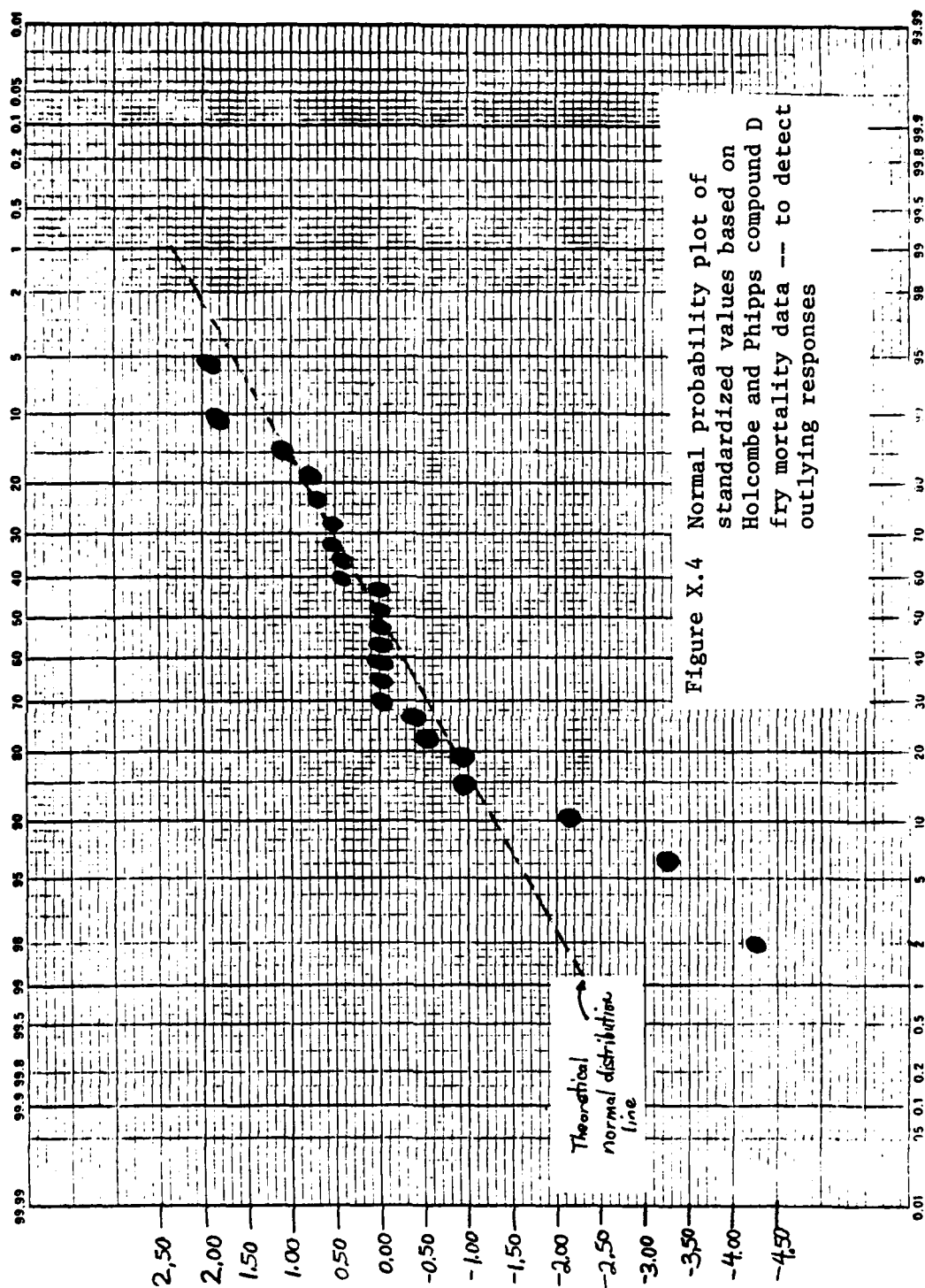


Figure X.4 Normal probability plot of standardized values based on Holcombe and Phipps compound D fry mortality data -- to detect outlying responses

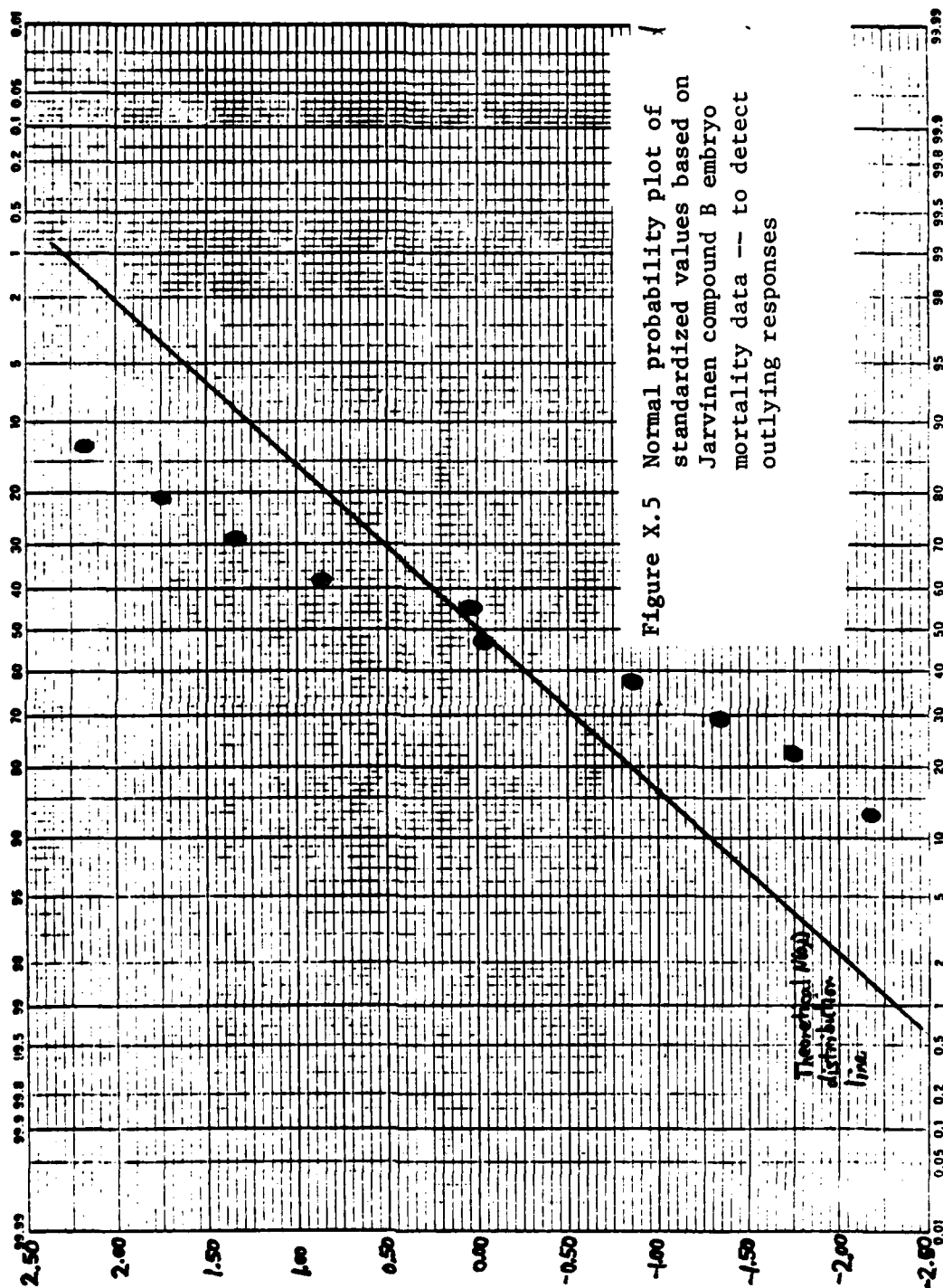


Figure X.5 Normal probability plot of standardized values based on Jarvinen compound B embryo mortality data -- to detect outlying responses

XI. TESTING FOR CONCENTRATION RELATED EFFECTS

A. Background

After we have completed preliminary graphical displays, tests for tank to tank heterogeneity, and outlier detection procedures we are ready to proceed to the main portion of the data analysis. This involves comparing responses across treatment groups to arrive at an inference about what constitutes an "acceptable" concentration. If no tank to tank heterogeneity is evident in the data then the original data may be pooled across tanks within groups and subsequent analyses carried out on a per fish basis or alternatively the data can first be adjusted to reflect the increased variation and the adjusted "data" can be pooled across tanks and analyzed on a per fish basis. As remarked earlier, we prefer the latter approach.

Before considering statistical procedures to determine acceptable concentrations, we must first define what is considered to be acceptable. According to the guidelines for early life stage tests, [8], "...A lower chronic endpoint is the highest tested concentration...which did not cause the occurrence (which was statistically significantly different from the control at the 95% level) of any specified adverse effect and below which no tested concentration caused such an occurrence...An upper chronic endpoint is the lowest tested concentration...which caused the occurrence (which was statistically significantly different from the control at the 95% level) of any specified adverse effect and above which all tested concentrations caused such an occurrence". We are thus interested in determining which concentrations yield (statistically) significantly different results than the control group. In a later section we will present an alternative notion of acceptable concentration.

Opinion: For the purpose of testing hypotheses concerning heterogeneity of response rates across groups or of constructing confidence intervals to compare treatment group and control group response rates, unless there is relatively strong statistical evidence of heterogeneity among tanks within groups (e.g. observed significance level less than 0.05 or 0.10) then act as if there is not heterogeneity of response from tank to tank within groups. Base subsequent tests and confidence intervals on the original (i.e. unadjusted) data, pooled across tanks within groups.

This suggestion reflects a conservative viewpoint with respect to the conclusions drawn from such subsequent analyses. Namely, suppose that such tank to tank variation within treatment groups exists but we do not detect it. Then we proceed as

if none exists. Thus the "true" variability of the test statistics that we use will exceed the assumed variability. Thus if we carry out a test at nominal level $\alpha = 0.05$, say, the "true" alpha level will in fact be greater than the nominal.

Inflating the actual α - level over the nominal level makes the test more prone to reject the hypothesis of equality of treatment and control groups than an actual α - level test. Thus if we err, it will be on the side of declaring a treatment group significantly different from the control group when it in fact is not. This is conservative.

Carrying out analyses on a per tank basis drastically reduces the degrees of freedom available for estimating variability of the test statistics, especially when there are very few tanks per treatment group. This diminishes the power of subsequent tests to compare treatment groups with control group, thus causing acceptance of the null hypothesis more often than necessary when it is false. That is, suppose there is little or no tank to tank heterogeneity within groups but we still carry out our analyses on a per tank basis "to be safe". Then the reduced power of the procedures based on per tank analyses will cause us to miss moderate differences between treatment groups and control groups that we might be able to detect if we were analyzing on a per fish basis. Thus we would be erring on the side of failing to detect departures from the control group response rates. This is unconservative.

It should be noted that exactly the opposite is true when we fit dose response curves to the data. Namely underestimation of variability results in an increased lower bound on safe concentration, which is unconservative. Thus for the purpose of fitting dose response curves perhaps we should adjust for tank to tank heterogeneity if the overall significance level is as great as $\alpha = .20$ or perhaps even $\alpha = .50$.

B. Chi Square Test of Homogeneity Across Treatment Groups

The most commonly used overall test for differences in mortality or abnormality rates across treatment groups is the chi square test for homogeneity. It is analagous to the shotgun analysis of variance F test for quantitative data. Since the preliminary tank to tank heterogeneity tests on the DeFoe and on the Holcombe and Phipps data sets were at most marginal, we use the original data. However the same test could be applied after adjusting the data.

Let p_1, p_2, \dots, p_I denote the response probabilities in the I (treatment and control) groups. We wish to test the hypothesis

$$H_0: p_1 = p_2 = \dots = p_I.$$

The chi square test is an overall test of this hypothesis (i.e. a shotgun test). It is based on the following statistic:

Let (X_{ij}, N_{ij}) denote the number of responses and the number of fish, respectively, in the j -th tank within the i -th treatment group

$$i = 1, \dots, I; j = 1, \dots, J.$$

Let

$$N_{i+} = \sum_{j=1}^J N_{ij}, X_{i+} = \sum_{j=1}^J X_{ij}, N_{++} = \sum_{i=1}^I \sum_{j=1}^J N_{ij}, X_{++} = \sum_{i=1}^I \sum_{j=1}^J X_{ij}. \text{ Then}$$

$$p_{++} \equiv \frac{X_{++}}{N_{++}} \text{ is the estimate of the common value of } p \text{ under } H_0.$$

$$\chi^2 \equiv \sum_{i=1}^I \frac{(X_{i+} - N_{i+} \hat{p}_{++})^2}{N_{i+} \hat{p}_{++} (1 - \hat{p}_{++})}$$

is the χ^2 test of H_0 . Under H_0 , $\chi^2 \sim \chi_{I-1}^2$. Since the chi square test is sensitive to all kinds of departures from H_0 , it is not tailor made to be sensitive to ordered alternatives, of the type most commonly encountered in toxicology. We will discuss this further later.

The chi square test is easy to carry out from a computational standpoint since many standard programs are available. For example the procedure PROC FREQ in the SAS statistical computing system [12] can be used to carry out this test. The program BMDP1F in the BMDP statistical computing system [9] can also be used for this purpose. Figure XI.1 illustrates output from SAS PROC FREQ to test for homogeneity of response rates across groups for the fry mortality data from the Holcombe and Phipps test on compound D. This test is based on data pooled across tanks within groups. Of course, the test rejects H_0 very strongly, as it should based on the appearance of the preliminary scatterplot.

We have also incorporated this test into our EXAX2 computer program [14]. We pool responses across tanks within groups and compute the chi square test. If expected frequencies within each cell exceed the cutoff, we evaluate the significance of chi square based on its asymptotic distribution under H_0 .

If any expected frequency is less than the cutoff we evaluate the exact small sample distribution of chi square, conditional on the margins, by enumeration as discussed in the section on the exact chi square program.

We illustrate this feature of EXAX2 with the DeFoe compound C and with the Holcombe and Phipps compound D data. We tested for heterogeneity of response rates across groups for the fry mortality and the embryo mortality data. The results of these tests are shown in Figures XI.2 - XI.5.

Figure XI.2 illustrates the chi square test of homogeneity across treatment groups for the embryo mortality data in the DeFoe compound C experiment. It will be recalled that no tank to tank heterogeneity within treatment groups was found by the preliminary test, and so the data have been pooled across the tanks within treatment groups. The first matrix displays the observed 2 x 6 table. The second matrix displays the expected (under H_0) frequencies. Since each expected frequency exceeds 5 (by a great deal in this example) the test is based on asymptotic theory.

NOTE THAT SINCE THIS IS A PRELIMINARY TEST WE SHOULD BE VERY LIBERAL IN DECIDING WHEN TO REJECT H_0 AND TO GO ON TO MORE DETAILLED COMPARISONS. THUS A LARGE α -VALUE e.g. $\alpha = .20$ or $\alpha = .25$ SHOULD BE USED. THIS ENHANCES THE SENSITIVITY OF THE TEST TO DETECT MODERATE DEPARTURES FROM H_0 .

We see from the bottom of Figure XI.2 that the observed significance level is $\alpha = 0.31$. Thus even by our liberal yardstick we see no statistical evidence of group to group differences in embryo mortality in the DeFoe data. This agrees with the appearance of the preliminary scatter plots.

The same test was carried out for the DeFoe compound C fry mortality data. The results are given in Figure XI.3. Again, all the expected frequencies exceed 5.0 and so the asymptotic theory is used. This time the chi square statistic is highly significant. Chi square = 182.79 with 5 d.f. Thus there is strong statistical evidence of group to group response rate differences in fry mortality. This again agrees with the appearance of the preliminary scatter plot.

Figures XI.4, XI.5 contain the results of the chi square tests of homogeneity across groups for the Holcombe and Phipps compound D embryo mortality and fry mortality data respectively. Again the data have been pooled across tanks within groups.

In both cases the expected frequencies exceed 5 and so asymptotic theory is used. We see for the embryo mortality data the test is nonsignificant even at the liberalized $\alpha = .20$. For the fry mortality data the test is very highly significant (chi square = 389.68 with 5 d.f.). Thus again there is no statistical evidence of group to group differences with respect to embryo mortality while there is strong statistical evidence of group to group differences with respect to fry mortality. This is in good agreement with the appearances of the preliminary scatterplots.

Figure XI.6, XI.7 contain the results of the chi square tests of homogeneity across groups for the Jarvinen compound B embryo mortality and fry mortality data respectively. The data have been pooled across tanks within groups. In both cases all the expected frequencies exceed 5 and so asymptotic theory is used.

For the fry mortality data the test is very highly significant, as was the case with the fry mortality data from the other experiments considered. It is quite clear that the last two treatment groups have substantially higher response rates than the first two groups.

In contrast to the cases for the DeFoe and Holcombe and Phipps data sets, there is some statistical evidence of group to group differences in embryo mortality rates.

We also saw strong statistical evidence in Sections VII and X of heterogeneity in response rates among tanks within groups. In Section IX we calculated an adjustment factor of $\hat{K} = 3.071$ for these responses, to account for the tank to tank heterogeneity. The effect of this adjustment on chi square is to adjust it downward by the factor \hat{K} . Thus with respect to the adjusted "responses", the observed chi square value becomes $10.71426 / \hat{K} = 10.71426 / 3.071 = 3.489$. The probability that a chi square random variate with 5 d.f. exceeds 3.489 is 0.625. Thus the tank to tank heterogeneity within groups accounts for the significant chi square across groups. Thus again there is no statistical evidence of variation in embryo mortality rate across groups.

C. One Sided Tests of Homogeneity Across Treatment Groups

The shotgun chi square test, although the most commonly used test of homogeneity of response rates, is not the most appropriate

test for application to toxicity data. The chi square test is an overall test which is not designed to be particularly sensitive to the one sided, monotone alternatives characteristic of dose response tests. More specialized tests have been designed to be more powerful against alternatives of this type.

Several tests of response rate homogeneity against ordered alternatives are discussed in the literature. Snedecor and Cochran [28] section 9.11 and Steel and Torrie [29], section 22.10 extract a single degree of freedom from the overall chi square test to test for linear regression in $2 \times K$ tables where the columns fall in a natural order. Scores, Z_j , are assigned (arbitrarily) to the columns to treat them as values on a continuous scale of measurement. The weighted linear regression coefficient of mortality probability on score Z_j is calculated and tested for significance. The major drawback of this method is the arbitrariness of the scores. See [28, 29] for details.

An alternative approach to the construction of one sided tests is by means of measures of association for ordered contingency tables. Such measures can be thought of as analogs for qualitative responses to correlation coefficients for quantitative responses. Goodman and Kruskal [30, 31] have derived and reported on a number of measures. Several commonly used measures of association are Kendall's tau b, Stuart's tau c, Goodman and Kruskal's gamma, just to name a few. For a given table each of these measures yields different numerical values and so it is not clear how to ascribe physical meaning to these values. However for each of the measures a value of zero means no monotone association between categories and positive or negative values mean positive or negative associations respectively. Thus a test of homogeneity of treatment group response rates that is sensitive to monotone, one sided alternatives can be constructed by testing the null hypothesis that these measures are zero against a one sided alternative. Brown and Benedetti [32] have calculated improved standard error estimates for the various measures that are appropriate studentizing factors to test the null hypothesis that these measures are zero. They show empirically that their new standard errors provide better approximations in small and moderate samples than do the older standard error estimates reported by Goodman and Kruskal [31]. Furthermore they show that a number of measures, each having different numerical values, result in identical "t ratios" when normalized by their respective Brown and Benedetti standard error estimates. This is desirable because we need consider just one "t ratio" rather than five. Proctor [33] shows that tests based on measures of association are in fact much more powerful against one sided, monotone alternatives than is the shotgun chi square test, as would be intuitively expected.

Agresti and Wackerly [15] also discuss one sided tests of homogeneity based on measures of association. They discuss Kendall's γ_b in some detail. They illustrate an instance of the increased sensitivity of such measure of association tests for detecting ordered departures from homogeneity with the following artificial example. Consider the 3 by 3 contingency table with ordered categories:

A \ B	B		
	High	Medium	Low
High	6	4	2
Medium	4	4	4
Low	2	4	6

There is clearly a positive trend in the table however the Fisher-Irwin test (exact) shows significance level $\alpha = .514$. This test would thus miss the trend. However Kendall's γ_b test (exact) is significant at $\alpha = .053$ and so would detect the trend.

Agresti and Wackerly also comment that the asymptotic normal approximation to the distribution of the sample estimate of γ_b may be quite poor for small sample sizes. They report that the observed significance level can be substantially greater than the nominal for small sample sizes; i.e. we reject H_0 when it is correct far more than the nominal proportion of times. This is at least in part due to the maximum likelihood estimate of standard error of γ_b having a negative bias. Based on Agresti and Wackerly's example, asymptotic normal distribution theory would be suspect at least for N below 50. Agresti and Wackerly suggest that an alternative, exact conditional test against ordered alternatives can be used, based on measures of association and enumeration of tables, when the sample sizes are too small to apply asymptotic normal theory.

The applicability of asymptotic distribution theory for the sample sizes and magnitudes of response proportions encountered in fish toxicity tests is a matter for detailed future study, probably by simulation. This is too involved for us to consider here. However as we use this test only on pooled data (original or adjusted) across tanks within groups, the sample sizes would be expected to be reasonably large (N in excess of 200) and so we utilize asymptotic theory for the remainder of this section.

It should be noted that ordinal measures of association assume that as one variable (e.g. concentration) increases the other variable (e.g. percent mortality or percent abnormality) either increases monotonically or decreases monotonically.

Nonmonotone relations (e.g. first increasing and then decreasing) can well result in small or even zero values of the measures. This is analogous to properties of correlation coefficients.

Goodman and Kruskal [30] define and discuss the properties of a number of measures of association for cross classifications. They include a section on measures for ordered categories (i.e. ordinal data). They propose a measure, γ , which is defined as follows:

Suppose two individuals are drawn at random from a population described jointly by two discrete, ordered categories.

Category 1: $i = 1, \dots, I$

Category 2: $j = 1, \dots, J$

In our fish toxicity examples $I = 2$ (e.g. live, dead), $J = C$ (number of groups)

Let (i, j) , (i', j') denote the (random) indices of these two individuals within the two categories. If there is an ordered correspondence between categories, we should see the same (or opposite) orderings of each of the categories, depending on direction of association.

$$\begin{aligned} \text{Let } \pi_s &\equiv P[(i > i' \text{ and } j > j') \text{ or } (i < i' \text{ and } j < j')]. \equiv P\{\text{same}\} \\ \pi_d &\equiv P[(i > i' \text{ and } j < j') \text{ or } (i < i' \text{ and } j > j')]. \equiv P\{\text{diff}\} \\ \pi_t &\equiv P[i = i' \text{ or } j = j']. \equiv P\{\text{tie}\} \end{aligned}$$

To avoid ambiguity they condition on the absence of ties. The conditional probability of like orders given no ties is $\pi_s / (1 - \pi_t)$. The conditional probability of unlike orders given no ties is $\pi_d / (1 - \pi_t)$. The difference of these two probabilities is defined as γ . Namely

Goodman and Kruskal's Gamma

$$\gamma = \frac{\pi_s - \pi_d}{1 - \pi_t} \equiv \frac{\pi_s - \pi_d}{\pi_s + \pi_t}$$

In the situation when the two categories are independent $\pi_s = \pi_d$ and so $\gamma = 0$. However the converse is not necessarily true (except in the 2×2 case).

The Kendall's γ_b and Stuart's γ_c measures are related to γ .
Let $m \equiv \min(I, J)$. Then

$$\gamma_c \equiv \frac{\Pi_s - \Pi_d}{(m - 1)/m}$$

This modification is made so that γ_c can nearly attain the absolute value 1 for nonsquare tables when the entire population lies on a longest diagonal of the table.

Kendall's γ_b is also related to γ . Namely, let p_{ij} denote a cell frequency,

$$p_{i.} \equiv \sum_{j=1}^J p_{ij}, \quad p_{.j} \equiv \sum_{i=1}^I p_{ij}. \quad \text{Then}$$

$$\gamma \equiv \frac{\Pi_s - \Pi_d}{1 - \Pi_t} = \frac{\Pi_s - \Pi_d}{1 - \sum_{i=1}^I p_{i.}^2 - \sum_{j=1}^J p_{.j}^2 + \sum_{i=1}^I \sum_{j=1}^J p_{ij}^2}$$

$$\gamma_b \equiv \frac{\Pi_s - \Pi_d}{\sqrt{\left[1 - \sum_{i=1}^I p_{i.}^2\right] \left[1 - \sum_{j=1}^J p_{.j}^2\right]}}$$

γ_b corrects for pairs of observations tied with respect to at least one of the categorizations and ranges between -1 and +1.

Somer's d (asymmetric measure)

$$d_{R|C} = \frac{\Pi_s - \Pi_d}{1 - \Pi_t + Y_o}$$

$$d_{C|R} = \frac{\pi_s - \pi_d}{1 - \pi_t + X_o}$$

Where $Y_o \equiv$ probability of tie in row only

$X_o \equiv$ probability of tie in column only

For all of these measures a zero value indicates a complete lack of a monotone relationship between the two variables (no association). A value of +1 indicate a perfect monotone increasing relationship (perfect agreement) and a - 1 indicates a perfect monotone decreasing relationship (perfect disagreement). It should be noted that lack of a monotone relationship is not the same as statistical independence. These measures will equal zero when there is dependence of a complicated form. However, when the variables are independent, the measures will equal zero. Kendall's γ_b differs from the others in that it can reach a value of ± 1 only for square tables, otherwise its maximum is lower.

Stuart's γ_c is an adjustment of γ_b that can attain value ± 1 for non square tables.

There is much discussion in the literature about which measures most realistically portray strengths of monotone relations. In general it is difficult to interpret the magnitudes of these measures in any physically meaningful fashion. However we will be using these measures of association only for tests of significance to detect departures from 0. For this application the situation simplifies considerably since Brown and Benedetti, page 311, show that basing a test of significance on γ_c , γ , γ_b , $d_{C|R}$, $d_{R|C}$ all lead to exactly the same test statistic, sample for sample. Thus we do not need to be concerned with differences among the values of the measures. That is, Brown and Benedetti, [125] have derived new estimates ASE_o of the asymptotic standard errors of the measures of association that are better than those given previously in the literature for testing the null hypothesis that the measure is zero. They report one set of standard errors to use for testing purposes and another set of standard errors to use for constructing confidence intervals. They show that

$$\frac{\hat{\gamma}}{ASE_o(\gamma)} \equiv \frac{\hat{\gamma}_b}{ASE_o(\hat{\gamma}_b)} \equiv \frac{\hat{\gamma}_c}{ASE_o(\hat{\gamma}_c)} \equiv \frac{\hat{d}_{R/C}}{ASE_o(\hat{d}_{R/C})} \equiv \frac{\hat{d}_{C/R}}{ASE_o(\hat{d}_{C/R})} \equiv T$$

which means that the five measures all give the same test of the null hypothesis of no (monotone) association.

Brown and Benedetti report a simulation study of the use of these T ratios to test the null hypothesis. They compared the T-values to the percentage points of the standard normal distribution. They concluded that

1. The ASE₀'s give empirical type I error rates closer to the nominal significance level and more consistent for different patterns of non-association than do previously reported standard error estimates.
2. For $N \geq 100$ the T - value can safely be compared to the percentile points of the standard normal distribution.
3. For $N \leq 50$ the distribution of the T - value seems to have heavier than normal tails, and they recommend comparing it to Student's t with approximate degrees of freedom (ADF) = $0.4N$.

Their T - ratios for testing the null hypothesis of nonassociation (i.e. monotone) and (asymptotic) standard errors appropriate for constructing confidence intervals on the measures of association are implemented in the BMDP [27] program BMDP1F, measures of association for two way frequency tables. (Note that BMDP1F was extensively rewritten and reissued in August, 1976. Thus only versions of this program dated after August, 1976 are based on the most up to date theory). We will illustrate the use of this program in this section, with both artificial and real data.

Proctor [33] has discussed the relative efficiencies of tests of association for ordered two way contingency tables and has compared these efficiencies with that of the chi square test. He reports that in most cases of ordered alternatives, the efficiencies of tests based on the measures of association are much greater than that of tests based on chi square. For one example of a 6 x 6 ordered contingency table constructed from an underlying bivariate normal distribution with correlation $\rho = 0.80$, the efficiencies of the tests of association based on measures of association relative to the chi square test were about 3.4. This means that for the chi square test to attain the same power against this alternative as a test based on the measures of association, it would need to be based on more than three times as many observations. In efficiency calculations based on other assumptions about the alternatives, the chi square procedure was consistently generally very much less efficient than test procedures based on measures of association.

To get some further feeling for the sensitivities to ordered alternatives of tests based on measures of association as compared with the chi square test we constructed several artificial sets of data having varying degrees of monotone trend in response probabilities. We tested the null hypothesis of no association between mortality level and treatment group using the one sided tests based on ordinal measures of association and the shot gun chi square test. Both of these tests can be carried out using the BMDP program [27] BMDP1F (versions subsequent to August, 1976). We should note that in these one sided tests we are looking for counterassociation. That is, the probability of being alive decreases as concentration group increases. We are thus testing for departures from 0 in a negative direction.

Figure XI.8 contains instructions for using program BMDP1F. Figures XI.9, 10 and 11 illustrate one sided and chi square tests on tables (based on artificial data) that exhibit linear trends of response probability with concentration group, but with differing slopes. They represent mild, moderate, and strong trends. In each case the one sided test based on measures of association reflects a much stronger association between categories than does the chi square test.

In conclusion, we see that the various ordinal measures of association provide equivalent tests of the null hypothesis of no association between mortality rate and concentration group. Further more all the tests are much more sensitive than the chi square test to alternatives of a monotonic nature.

Appendix A5, pp 778-792, of the 1979 BMDP manual [27] and Brown and Benedetti [32] are helpful in interpreting the output from BMDP1F. Brown and Benedetti calculate two asymptotic standard errors for each measure, ASE_0 , ASE_1 . ASE_1 is derived assuming the alternative hypothesis is true; i.e. the measure is not zero. It is obtained by the method of Goodman and Kruskal [31] and is appropriate for setting confidence limits on the measures for large samples. Brown and Benedetti discuss the use of ASE_1 in computing confidence limits and power in an unpublished technical report that is available from the Health Sciences Computing Facility at UCLA.

ASE_0 is computed under the null hypothesis that the measure is zero. It was derived by Brown and Benedetti in the 1977 paper cited above.

The T-value for each measure is the ratio of each measure to its ASE_0 . Brown and Benedetti report, based on simulation studies, that the use of ASE_0 in the denominator of the T-statistic rather than ASE_1 or other suggestions made in the literature gives superior results in that the attained type I error

rates are closer to nominal and are more consistent for differing patterns of probabilities.

To illustrate the use of these one sided tests of association vs ordered alternatives as compared with the chi square test of significance, we ran the BMDP program, BMDP1F, on mortality data from the DeFoe compound C test and from the Holcombe and Phipps compound D tests. The results are shown in Figures XI.12-15, both for embryo mortality data and for fry mortality data. Qualitatively there is no difference, in these data, in the conclusions arrived at by each procedure. The relationships between concentration and percent response are so strong in the fry mortality data that the observed significance levels are 0 to many decimal places. For the embryo mortality data, neither procedure reveals a statistically significant relation between concentration and percent mortality. Since the observed mortality in the DeFoe embryo data is smaller at the higher concentrations than at the lower, the measures of association are positive and the observed significance level is higher for the one sided test than for the chi square test. We almost have a significant trend in the wrong direction! Why? It may be due to the outlier tank in group 2.

In conclusion it should be remarked that any overall test for concentration related effects is just a screening device. It merely states whether there is any statistical evidence of concentration related effects but does not provide any indication of which treatment groups have responses that differ from the control group. That is the role of multiple comparisons. The overall test is intended to screen out those data sets for which multiple comparisons would be a futile exercise because no differences exist. In this regard it should be noted that since the overall test is just a preliminary procedure, it makes sense to use a very liberal α -level, like $\alpha = .20$ or perhaps even $\alpha = .50$. This improves the sensitivity of the test to detect marginal effects, but at the expense of an increased false rejection rate. However such false rejections of the null hypothesis will be detected later in the multiple comparison phase.

CTGL	TRT							
FREQUENCY	FRY MORTALITY							
EXPECTED								
CELL CHI2								
COL PCT	1	2	3	4	5	6	TOTAL	
DEAD	6	8	8	13	79	100	214	
	35.7	35.7	35.7	35.7	35.7	35.7		
	24.7	21.5	21.5	14.4	52.6	116.0		
	6.00	8.00	8.00	13.00	79.00	100.00		
LIVE	94	92	92	87	21	0	366	
	64.3	64.3	64.3	64.3	64.3	64.3		
	13.7	11.9	11.9	8.0	29.2	64.3		
	94.00	92.00	92.00	87.00	21.00	0.00		
TOTAL	100	100	100	100	100	100	600	

STATISTICS FOR 2-WAY TABLES

CHI-SQUARE	389.676	DF=	5	PROB=0.0001
PHI	0.806			
CONTINGENCY COEFFICIENT	0.627			
CRAMER'S V	0.806			
LIKELIHOOD RATIO CHISQUARE	444.801	DF=	5	PROB=0.0001

Figure XI.1 SAS PROC FREQ output from chi square test for homogeneity across groups applied to fry mortality data from Holcombe and Phipps test on Compound D

DEFOE COMPONDS C DEFENSE MORTALITY

CONTINGENCY TABLE

GROUP 1

	ALIVE	DEAD	TOTAL
COL 1	59.000	41.000	100.000
COL 2	49.000	51.000	100.000
COL 3	63.000	37.000	100.000
COL 4	63.000	37.000	100.000
COL 5	61.000	39.000	100.000
COL 6	62.000	38.000	100.000

TOTAL ALIVE = 357.000 , TOTAL DEAD = 243.000 , GRAND TOTAL = 600.000

EXPECTED FREQUENCIES

XSQDIFF = 5.95407 CUTOFF = 5.00000

	ALIVE	DEAD	TOTAL
COL 1	59.500	40.500	100.000
COL 2	59.500	40.500	100.000
COL 3	59.500	40.500	100.000
COL 4	59.500	40.500	100.000
COL 5	59.500	40.500	100.000
COL 6	59.500	40.500	100.000

TOTAL ALIVE = 357.000 , TOTAL DEAD = 243.000 , GRAND TOTAL = 600.000

1. INDEPENDENT TESTS COMBINED

ASYMPTOTIC CHI SQUARE USED

DESERVED CHANCE = 5.95407

DESERVED CHANCE LEVEL = 0.2102677745

Figure XI.2 EXAX2 output -- chi square test of homogeneity across groups

DEFOE COMPOUND C FRY MORTALITY

CONTINGENCY TABLE

GROUP	1	2	3	4	5	6	TOTAL	ALIVE	DEAD	TOTAL
CUL	1	2	3	4	5	6	140.000	140.000	0.0	140.000
CUL	2	3	4	5	6	7	50.000	50.000	0.0	50.000
CUL	3	4	5	6	7	8	40.000	40.000	0.0	40.000
CUL	4	5	6	7	8	9	40.000	40.000	0.0	40.000
CUL	5	6	7	8	9	10	40.000	40.000	0.0	40.000
CUL	6	7	8	9	10	11	40.000	40.000	0.0	40.000
TOTAL	ALIVE	DEAD	TOTAL	ALIVE	DEAD	TOTAL	50.000	50.000	0.0	100.000

GROUP	1	2	3	4	5	6	TOTAL	ALIVE	DEAD	TOTAL
CUL	1	2	3	4	5	6	100.000	100.000	0.0	100.000
CUL	2	3	4	5	6	7	50.000	50.000	0.0	50.000
CUL	3	4	5	6	7	8	40.000	40.000	0.0	40.000
CUL	4	5	6	7	8	9	40.000	40.000	0.0	40.000
CUL	5	6	7	8	9	10	40.000	40.000	0.0	40.000
CUL	6	7	8	9	10	11	40.000	40.000	0.0	40.000
TOTAL	ALIVE	DEAD	TOTAL	ALIVE	DEAD	TOTAL	50.000	50.000	0.0	100.000

1 INDEXED NO ONE TESTS COMPLETED

ASYMPTOTIC CHI SQUARED USED

RESERVED FOR FUTURE USE L V L = 100.000000 0.0

Figure XI.3 --EXAX2 output-- chi square test of homogeneity across groups

*** THIS TABLE IS FOR INFORMATION ONLY ***

HOLCOMBE AND PHIPPS COMPOUND D EMBRYO MORTALITY

*** THIS TABLE IS FOR INFORMATION ONLY ***

GROUP 1		CONTINGENCY TABLE			
		ALIVE	DEAD	TOTAL	
COL 1	1	130.000	70.000	200.000	
COL 2	2	130.000	70.000	200.000	
COL 3	3	137.000	63.000	200.000	
COL 4	4	127.000	73.000	200.000	
COL 5	5	133.000	67.000	200.000	
COL 6	6	124.000	76.000	200.000	
TOTAL		ALIVE = 776.000	DEAD = 424.000	GRAND TOTAL = 1200.000	
		EXPECTED FREQUENCIES			
		ALIVE	DEAD	TOTAL	
COL 1	1	129.333	70.667	200.000	
COL 2	2	129.333	70.667	200.000	
COL 3	3	129.333	70.667	200.000	
COL 4	4	129.333	70.667	200.000	
COL 5	5	129.333	70.667	200.000	
COL 6	6	129.333	70.667	200.000	
TOTAL		ALIVE = 776.000	DEAD = 424.000	GRAND TOTAL = 1200.000	
		XSQORS = 3.39914 CUTOFF = 5.00000			

1 INDEPENDENT TESTS COMBINED
 ASYMPTOTIC CHI SQUARE USED
 OBSERVED CHI SQUARE = 3.39914
 OBSERVED SIGNIFICANCE LEVEL = 0.02700756

Figure XI.4 --EXAX2 output-- chi square test of homogeneity across groups

HOLCOMBE AND PHIPS COMPOUND D FRY MORTALITY

CONTINGENCY TABLE

GROUP 1

	ALIVE	DEAD	TOTAL
CUL 1	94.000	6.000	100.000
CUL 2	92.000	8.000	100.000
CUL 3	92.000	8.000	100.000
CUL 4	87.000	13.000	100.000
CUL 5	21.000	79.000	100.000
CUL 6	0.0	100.000	100.000

TOTAL ALIVE = 396.000 , TOTAL DEAD = 214.000 , GRAND TOTAL = 610.000

EXPECTED FREQUENCIES

	ALIVE	DEAD	TOTAL
CUL 1	64.333	35.667	100.000
CUL 2	64.333	35.667	100.000
CUL 3	64.333	35.667	100.000
CUL 4	64.333	35.667	100.000
CUL 5	64.333	35.667	100.000
CUL 6	64.333	35.667	100.000

TOTAL ALIVE = 386.000 , TOTAL DEAD = 214.000 , GRAND TOTAL = 600.000

1 INDEPENDENT TESTS COMBINED

ASYMPTOTIC CHI-SQUARED USED

38.1150 CHI-SQUARED = 38.1150

38.1150 SIGNIFICANCE = 0.0

Figure XI.5 --EXAX2 output-- chi square test of homogeneity across groups

JARVINEN COMPOUND B ENDOTOXICITY

GROUP 1		CONTINGENCY TABLE			
		ALIVE	DEAD	TOTAL	
TRT	1	37.000	14.000	51.000	
TRT	2	101.000	8.000	109.000	
TRT	3	91.000	9.000	100.000	
TRT	4	96.000	2.000	98.000	
TRT	5	99.000	6.000	105.000	
TRT	6	93.000	8.000	101.000	
TOTAL		567.000	47.000	614.000	
		GRAND TOTAL = 614.000			
		EXPECTED FREQUENCIES			
		ALIVE	DEAD	TOTAL	
TRT	1	93.269	7.731	101.000	
TRT	2	100.656	8.344	109.000	
TRT	3	92.345	7.655	100.000	
TRT	4	90.498	7.502	98.000	
TRT	5	95.463	9.537	105.000	
TRT	6	93.269	7.731	101.000	
TOTAL		567.000	47.000	614.000	
		GRAND TOTAL = 614.000			
		CUTOFF = 5.00000			

1 INDEPENDENT TESTS COMPILED
ASYMPTOTIC CHI-SQUARED USED
OBSERVED CHI-SQUARED = 10.71426
OBSERVED SIGNIFICANCE LEVEL = 0.0573491071

Figure XI.6 --EXAX2 output-- chi square test of homogeneity across groups

JARVINEN COMPOUND B FAT MORTALITY

GROUP 1 CUNTINGINCY TABLE

	ALIVE	DEAD	TOTAL
TRI 1	30.000	0.0	30.000
TRI 2	29.000	0.0	29.000
TRI 3	30.000	0.0	30.000
TRI 4	30.000	0.0	30.000
TRI 5	14.000	16.000	30.000
TRI 6	0.0	30.000	30.000
TOTAL	ALIVE = 133.000 , TOTAL	DEAD = 46.000 , GRAND TOTAL =	179.000

CUTOFF = 5.00000

EXPECTED FREQUENCIES

	ALIVE	DEAD	TOTAL
TRI 1	22.291	7.709	30.000
TRI 2	21.547	7.453	29.000
TRI 3	22.291	7.709	30.000
TRI 4	22.291	7.709	30.000
TRI 5	22.291	7.709	30.000
TRI 6	22.291	7.709	30.000
TOTAL	ALIVE = 133.000 , TOTAL	DEAD = 46.000 , GRAND TOTAL =	179.000

1 INDEPENDENT TESTS COMBINED

ASYMPTOTIC CHI SQUARED USED
 OBSERVED CHI SQUARED = 125.89580
 OBSERVED SIGNIFICANCE LEVEL = 0.0

Figure XI.7 --EXAX2 output-- chi square test of homogeneity across groups

BMDP1F - TWO-WAY FREQUENCY TABLES--MEASURES OF ASSOCIATION PROGRAM REVISED NOVEMBER, 1978
 HEALTH SCIENCES COMPUTING FACILITY
 UNIVERSITY OF CALIFORNIA, LOS ANGELES
 COPYRIGHT (C) 1977, THE REGENTS OF THE UNIVERSITY OF CALIFORNIA.

----- EXTENSIVE CHANGES HAVE BEEN MADE TO THIS PROGRAM.
 BE SURE TO USE 1977 MANUAL.

----- WHEN LESS CATEGORY NAMES ARE DEFINED THAN THERE ARE CATEGORIES, WHICH IS TREATED AS A MISSING VALUE.
 THE UNNAMED CATEGORIES ARE COLLAPSED INTO A SINGLE CATEGORY.

----- DEFINE OPTION DOES NOT FUNCTION.

PROGRAM CONTROL INFORMATION

/PROBLEM TITLE IS 'SENSITIVITY TESTING OF ORDINAL MEASURES'.
 /INPUT VARIABLES ARE 2.
 /VARIABLE FORMAT IS '16F5.0'.
 /TABLE NAMES ARE CONC, MORT.
 /CATEGORY ROW IS MORT.
 /STATISTICS COLUMN IS CONC.
 /END NAMES (2) ARE ALIVE, DEAD.
 /END GAMMA.

PROBLEM TITLE.....SENSITIVITY TESTING OF ORDINAL MEASURES

NUMBER OF VARIABLES TO READ IN: 2
 NUMBER OF VARIABLES ADDED BY TRANSFORMATIONS: 0
 TOTAL NUMBER OF VARIABLES: 2
 NUMBER OF CASES TO READ IN: 1000000
 CASE LABELING VARIABLES: 0
 LIMITS AND MISSING VALUE CHECKED BEFORE TRANSFORMATIONS: 0
 BLANKS ARE: 0
 INPUT UNIT NUMBER: 5
 REWIND INPUT UNIT PRIOR TO READING: 0
 DATA: 0
 NO
 INPUT FORMAT
 16F5.0

VARIABLES TO BE USED 2 MORT
 1 CONC

Figure XI.8 Input information required for program BMDP1F

TABLE NO. 4 MORT (VAR 2) VS CONC (VAR 1)
CELL FREQUENCY COUNTS

		CONC (VAR 1)							
		EQ./EQ.	1.00	2.00	3.00	4.00	5.00	6.00	TOTAL
MORT (VAR 2)	DEAD	1.00	6	7	8	9	10	11	51
	ALIVE	2.00	94	93	92	91	90	89	549
	TOTAL		100	100	100	100	100	100	600

MINIMUM NONZERO EXPECTED VALUE IS 8.500
ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	PEARSON CHISQUARE	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.	
		2.250	5	0.8136					
STATISTIC		VALUE	ASE1	T-VALUE	DEP.	VALUE	ASE1	T-VALUE	DEP.
GAMMA		-0.150	0.098	-1.504		-0.039	0.026	-1.504	
SUMER'S D		-0.125	0.082	-1.504	1	-0.054	0.035	-1.504	2
					SOMER'S D	-0.023	0.016	-1.504	2

Observed Significance Levels

1. Chi Square Test $\alpha = 0.81$

2. One Sided Measure of Association Test $\alpha = \Phi(-1.504) = 0.066$

Figure XI.9 BMDP1F output from one sided measure of association test on artificial data

TABLE NO. 2 MORT (VAR 2) VS CONC (VAR 1)

CELL FREQUENCY COUNTS

		EQ./EQ.	1.00	2.00	3.00	4.00	5.00	6.00	TOTAL
MORT (VAR 2)	DEAD	1.00	6	8	10	12	14	16	66
	ALIVE	2.00	94	92	90	88	86	84	534
	TOTAL		100	100	100	100	100	100	600

MINIMUM NONZERO EXPECTED VALUE IS 11.000
 ----- ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	7.150	5	0.2097				
STATISTIC				STATISTIC			
GAMMA	-0.238			STUART'S TAU-C	-0.078		
SOMER'S D	-0.199			KENDALL'S TAU-B	-0.096		
				SOMER'S D	-0.047		
				T-VALUE DEP.	-2.699		
				ASE1	0.029		
				0.035			
				0.017			
				T-VALUE DEP.	-2.699		
				0.035			
				0.017			

Observed Significance Levels

1. Chi Square Test $\alpha = 0.21$
2. One Sided Measure of Association Test $\alpha = \Phi(-2.699) = 0.0085$

Figure XI.10 BMDP1F output from one sided measure of association test on artificial data

Defoe Compound C Embryo Mortality Data

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN

TABLE NO. 1 MORT VAR 2) VS TRGRP (VAR 1)

CELL FREQUENCY COUNTS

		TRGRP (VAR 1)							
		EQ./EQ.	1.00	2.00	3.00	4.00	5.00	6.00	TOTAL
PORT (VAR 2)	DEAD	1.00	41	51	37	37	39	38	243
	ALIVE	2.00	59	49	63	63	61	62	357
	TOTAL		100	100	100	100	100	100	600

MINIMUM NONZERO EXPECTED VALUE IS 40.500
USING THE ORIGINAL CELL FREQUENCIES
ALL STATISTICS ARE COMPUTED EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC		VALUE	D.F.	PROB.	STATISTIC		VALUE	D.F.	PROB.
PEARSON CHI SQUARE		5.955	5	0.3106					
STATISTIC	GAMMA	VALUE	ASE1	T-VALUE	STATISTIC	TAU-C	VALUE	ASE1	T-VALUE
		0.070	0.057	1.241			0.057	0.046	1.241
SOMER'S D		0.059	0.047	1.241	SOMER'S D		0.045	0.034	1.241
							0.034	0.029	1.241

Observed Significance Levels

1. Chi Square Test $\alpha=0.31$
2. One Sided Measure of Association Test $\alpha=0.89$

Figure XI.12 BMDP1F output on real data

Defoe Compound C Fry Mortality Data

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 3)

TABLE NO. 1 MORT (VAR 2) VS TRTGRP (VAR 1)

CELL FREQUENCY COUNTS

	TRTGRP (VAR 1)						TOTAL
	EO./EQ.	1.00	2.00	3.00	4.00	5.00	
MORT 2) DEAD (VAR 1)	1.00	0	0	2	1	0	3
	2.00	40	40	38	40	31	189
TOTAL		40	40	40	41	40	241

MINIMUM NONZERO EXPECTED VALUE IS 0.63
ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	182.785	5	0.0	STATISTIC			
GAMMA	-0.963			ASE1	0.022	-10.615	
SUMER'S D	-0.925	0.029	-10.615	TAU-C	-0.626		
				KENDALL'S TAU-B	-0.589		
				SOMER'S D	-0.375		
				T-VALUE	0.059	-10.615	
				DEP.	0.032	-10.615	
					0.035	-10.615	2

Observed Significance Levels

1. Chi Square Test

$\alpha = 0.000$

2. One Sided Measure of Association Test

$\alpha = 0.000$

Figure XI.13 BMPDIF output on real data

HOLCOMBE AND PHIPPS COMPOUND D EMBRYO MORTALITY DATA

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (VAR 3)

TABLE NO. 1 MORT (VAR 2) VS TATGRP (VAR 1)

CELL FREQUENCY COUNTS

		TRTGRP (VAR 1)							
		EQ./Eq.	1.00	2.00	3.00	4.00	5.00	6.00	TOTAL
PORT (VAR 2)	DEAD	1.00	70	70	63	78	67	76	424
	ALIVE	2.00	130	130	137	122	133	124	776
	TOTAL		200	200	200	200	200	200	1200

MINIMUM NONZERO EXPECTED VALUE IS 70.667
----- ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE	3.399	5	0.6387				
GAMMA	-0.026			STUART'S TAU-C	-0.020		
SOMER'S D	-0.022			KENDALL'S TAU-B	-0.016		
				SOMER'S D	-0.012		
				ASE1	0.032		
				T-VALUE	-0.634		
				DEP.	0.026		
					0.019		
					-0.634		
					2		

Observed Significance Levels

1. Chi Square Test

$\alpha = 0.64$

2. One Sided Measure of Association Test

$\alpha = 0.26$

Figure XI.14 BMDPIF output on real data

HOLCOMBE AND PHELPS COMPOUND D FRY MORTALITY DATA

THE ANALYSIS IS OF THE DATA
USING THE FREQUENCY COUNTS IN FREQ (IVAR 3)

TABLE NO. 1 MURT (IVAR 2) VS TRIGRP (IVAR 1)

CELL FREQUENCY COUNTS

		TRIGRP (IVAR 1)					
		EQ./EQ.	1.00	2.00	3.00	4.00	5.00 6.00 TOTAL
MURT (IVAR 2)	DEAD	1.00	6	8	8	13	79 214
	ALIVE	2.00	94	92	92	87	100 386
TOTAL			100	100	100	100	100 600

MINIMUM NONZERO EXPECTED VALUE IS 35.667
ALL STATISTICS ARE COMPUTED USING THE ORIGINAL CELL FREQUENCIES
EXCLUDING ROWS AND/OR COLUMNS WHICH ARE ZERO.

STATISTIC	PEARSON CHISQUARE	VALUE	D.F.	PROB.	STATISTIC	VALUE	D.F.	PROB.
PEARSON CHISQUARE		389.676	5	0.0				
GAMMA		VALUE	ASE1	T-VALUE	DEP.	VALUE	ASE1	T-VALUE
SOMER'S D		-0.833	0.025	-24.242	1	-0.764	0.032	-24.242
		-0.833	0.020	-25.242	1	-0.618	0.022	-24.242
						-0.459	0.019	-24.242

Observed Significance Levels

1. Chi Square Test $\alpha = 0.000$

2. One Sided Measure of Association Test $\alpha = 0.000$

Figure XI.15 BMDP1F output on real data

XII. TREATMENT GROUP VS CONTROL GROUP PAIRWISE MULTIPLE COMPARISON PROCEDURES

If the overall test rejects the hypothesis of no concentration related effects we must determine which treatment group response rates differ from the control group rate. A number of procedures can be used for such inferences, some based on hypothesis testing and some based on confidence interval estimation. In this section we consider several approaches based on tests of hypotheses. The discussion is by no means exhaustive. In the following section we discuss confidence interval procedures.

A common approach to multiple comparisons on qualitative response rate data such as mortality rates is to carry out an arc sine normalizing transformation on the observed response rate within each group and then compare each treatment group with the control using Dunnett's or Williams' procedures [34, 35, 36, 37]. Such procedures are based on asymptotic theory whose validity is questionable if there are a number of small expected frequencies.

An alternative multiple testing approach is to carry out a succession of 2×2 contingency table tests of homogeneity between each treatment group and the control group, based on Fisher's exact test [13] or on asymptotic theory depending on expected frequencies. Our EXAX2 program will do this. A treatment group is said to be (statistically) significantly different from the control group at e.g. the $\alpha = 0.05$ level if the pairwise test rejects the null hypothesis after adjusting for simultaneity by Bonferroni's method. (i.e. If we perform five pairwise comparisons and wish to guarantee an overall $\alpha = 0.05$ type one error level then each individual comparison must be made at the $\alpha/5 = 0.01$ level). Note that this approach does not impose any monotonicity structure on the response rates and so may not be most sensitive to detect small to moderate effects.

Dunnett [34, 35] presents a procedure for multiple comparison of each of the treatment group responses with the control group response, controlling the overall error rate for all comparisons. His procedure is derived for quantitative responses, assumed to have equal variability. He assumes equal replication among the treatment groups with equal or possibly greater replication of the control group. We might apply this procedure to qualitative response data from toxicity tests after performing an arc sine variance stabilizing relation on the observed responses.

A problem with the application of Dunnett's procedure to the analysis of data from toxicity tests is that it does not take full account of the structure of the problem. Namely the various treatment groups correspond to increasing toxicant levels. One might therefore assume a monotone (increasing or decreasing) response level with increasing group number. Since Dunnett did not build such a monotonicity assumption into his procedure, it loses some sensitivity.

Williams [36, 37] assumes a monotone response function. He estimates the treatment and control group response rates under the monotonicity restraint and uses these estimates for treatment group - control group comparisons. See Williams [36] for details. Chew [38, pp.26-27] briefly describes Williams' method and presents tables for its implementation. Williams [36] assumes equal replication for all concentrations (including the control). He extends this procedure [37] to accommodate increased replication in the control group, two sided tests, and modifications to account for unequal replication among the treatment groups. As with Dunnett's procedure, we can apply Williams method to qualitative response data after carrying out an arc sine transformation.

We illustrate Williams' method with several examples based on results from fish toxicity tests. Consider first the fry mortality data from the Holcombe and Phipps test on compound D. From the preliminary scatterplot in Figure VI.6 and the overall tests of significance in Section XI, it is quite evident that fry mortality increases with increasing toxicant level. We wish to determine here which treatment groups exhibit significantly greater fry mortality rates than the control group. As the result of the within groups heterogeneity test was marginal ($\alpha = 0.14$, see section VIII) we do not adjust the data prior to carrying out Williams' procedure.

The basic and transformed responses, pooled across tanks within groups are:

Group (i)	1	2	3	4	5	6
Sample Size (n_i)	100	100	100	100	100	100
Response Rate (\hat{p}_i)	0.06	0.08	0.08	0.13	0.79	1.00
$2\text{Arc Sin } \sqrt{\hat{p}_i} \equiv \hat{\mu}_i$	0.495	0.574	0.574	0.738	2.190	3.142

Since these estimates are already in monotone sequence, they do not need to be modified. We declare the group i response rate to be significantly different from the control rate if

$$\hat{\mu}_i - \hat{\mu}_1 > \bar{t}(2/n)^{1/2}$$

The factor \bar{t} can be obtained from Williams' tables corresponding to the 5% or the 1% significance level. The yardstick $\bar{t}(2/n)^{1/2}$ is based on the assumption that the variance of $2 \text{ arc sin } \sqrt{\hat{p}}$ is $1/n$. In our example $n = 100$ and $\bar{t} = 1.756$ (corresponding to 5 treatment groups and $\alpha = 0.05$). Thus the response in group i is declared to differ significantly from

the control group response if $\hat{\mu}_1 > 0.495 + 1.756 (2/100)^{1/2} = 0.743$. Groups 5 and 6 differ significantly from the control and group 4 is just on the borderline.

We now examine the effect on the outcome of this procedure of applying an adjustment for tank to tank heterogeneity. From Section IX B we see that this factor is $\bar{K} = 1.337$ for the above data. Thus the "effective" sample size per group is $100/1.337 = 74.79$ and the decision point for Williams' procedure becomes $0.495 + 1.756(2/74.79)^{1/2} = 0.782$. Group 4 is no longer borderline.

We now apply this same procedure to the embryo mortality data from the Jarvinen test on compound B. The result of the within groups heterogeneity test was highly significant ($\alpha = 0.005$) and so we first adjust the data prior to carrying out Williams' procedure. From Section IXB the adjustment factor is $\hat{K} = 3.071$. The basic and transformed responses along with effective sample sizes, pooled across tanks within groups are:

Group (i)	1	2	3	4	5	6
Effective Sample size (n_i/\hat{K})	32.89	35.49	32.56	31.91	34.19	32.89
Response Rate (\hat{p}_i)	0.139	0.073	0.090	0.020	0.057	0.079
$2 \text{ Arc Sin } \sqrt{\hat{p}_i} \equiv \hat{\mu}_i$	0.764	0.547	0.609	0.284	0.482	0.570

For the sake of simplicity we will utilize an average sample size of 33.32 within each group, but the calculation could alternatively be carried out based on the individual group sample sizes. Since $\{\hat{\mu}_i\}$ are not in monotone sequence we must first modify them by an averaging process discussed in Williams [36] or in Chew [38] until the resulting estimates satisfy the monotonicity constraint. We obtain 0.537, 0.537, 0.537, 0.537, 0.537, 0.570. We declare group i significantly greater than the control group if $\hat{\mu}_i > 0.764 + 1.756 (2/33.32)^{1/2} = 1.194$. Obviously no treatment groups have significantly greater response than the control group. (Interestingly if we carry out Williams' procedure on these data to look for a monotone decreasing trend in response rate, we arrive at the same conclusion. That is, no group has significantly lower response rate than the control group).

Dunnett's and Williams' procedures are based on asymptotic theory. If the response frequencies do not justify the use of

asymptotic theory we can carry out a succession of exact, small sample 2 x 2 treatment-control comparisons by means of Fisher's exact test, adjusting for simultaneity by Bonferroni's method. Consider for example the comparison of treatment group 4 with the control group for the fry mortality data from Holcombe and Phipp's test on compound D. We have the following 2 x 2 table.

	CONTROL	GROUP 4	
DEAD	6	13	19
LIVE	94	87	181
	100	100	200

Here, in the notation of Lieberman and Owen,

$$\begin{aligned} k &= 19 \\ n &= 100 \\ N &= 200 \\ x &= 6 \end{aligned}$$

Interpolating in the Lieberman and Owen tables [26]

between $N = 100$ and $N = \infty$ we have

$N = 100$	$P(X \leq 6) = .062$	$1/N = .01$
$N = \infty$	$P(X \leq 6) = .0835$	$1/N = 0$
$N = 200$	$P(X \leq 6) = ?$	$1/N = .005$

$$\text{Thus } P(X \leq 6) = 1/2(.0835 + .062) = .073$$

Thus this table is significant at the .07 level (not accounting for simultaneity).

This exact test procedure is thus seen to be somewhat less sensitive than Williams' procedure applied to the same data. This is understandable since it does not incorporate the monotonicity structure of the response rates.

XIII. CONFIDENCE INTERVAL PROCEDURES FOR COMPARISON OF TREATMENT GROUP AND CONTROL GROUP RESPONSE RATES

A. Introduction

We have previously considered overall tests of hypotheses to compare response rates in replicate tanks within treatment groups and to compare response rates across treatment groups. In this section we consider procedures for constructing confidence intervals to compare response rates in the treatment groups with that in the control group on a pairwise basis.

It is well known that hypothesis testing procedures are somewhat limited in their conclusions. They merely state whether the null hypothesis was accepted or rejected but give no indication of the extent of the effect. Thus we have no idea of the biological significance of the effect (as opposed to its statistical significance). The rejection or nonrejection of a null hypothesis is often more a result of sample size than of the biological importance of the effect. The determination of acceptable concentrations should be based on what are biologically significant effects rather than on the power function of a hypothesis testing procedure.

Confidence intervals are more informative than tests of hypothesis. The widths of the confidence intervals indicate the degree of precision in the data concerning the estimates of the quantities of interest in our inferences. Narrow confidence intervals signify precise inferences while wide confidence intervals signify imprecise inferences.

In the discussion in this section we consider the case of no tank to tank heterogeneity within groups. Thus we pool responses across tanks within groups to arrive at average response rates within groups. The presence of tank to tank heterogeneity can be accounted for by

1. Fitting a model which explicitly accounts for heterogeneity of response rates across tanks -- for example the beta binomial extension of the binomial model, the negative binomial extension of the Poisson model, or a variance components extension of a fixed effects analysis of variance model for quantitative responses.
2. By carrying out analyses on a per tank basis rather than on a per fish basis. This approach is conservative and greatly diminishes the number of degrees of freedom available for error estimation.
3. By adjusting the data to account for the extent of tank to tank variation. Namely tank to tank variation can be

regarded as correlated responses within tanks, generally positively correlated. Thus the variability of the average responses within tanks is greater than would be the case if the responses were independent within tanks. Such reduction in variation can be simply accounted for by reducing the "effective" sample size within tanks to a lesser value and then ignoring the within tank correlation and proceeding with binomial based procedures or the like. The reduction in "effective" sample size reduces the precisions of the estimates and test statistics just as does correlation effects.

The procedures discussed in this section, although based on binomial theory, can be used in conjunction with adjustment method 3. Thus they are also relevant in the case when tank to tank heterogeneity exists.

Consider the Holcombe and Phipps compound D fry mortality data. We wish to compare the response rate in treatment group 4 with that in treatment group 1 (the control group). The basic data, pooled across tanks within groups, is

	CONTROL	GROUP 4	
DEAD	6	13	19
LIVE	94	87	181
	100	100	200

Fisher's exact test (without simultaneity adjustment) says that p_4 is "significantly" greater than p_1 at the $\alpha = 0.07$ level. However a significance statement such as this says nothing about the magnitude of p_4/p_1 . Estimating the value of this ratio is important for assessing whether there is a biologically significant increase in mortality between the control group and group four. Confidence interval procedures enable us to estimate p_4/p_1 and determine the precision of our estimate as well determine whether p_4 is (statistically) significantly greater than p_1 .

There are three approaches to the construction of confidence intervals in the case of quantal response data.

- Large sample normal theory confidence intervals.
- Exact, small sample confidence intervals based on the noncentral distribution of the 2 x 2 contingency table, conditional on the margins. (See Thomas [39]) for the theory and the algorithm.

We have implemented this algorithm in EXAX2 [14].

- Approximate confidence intervals based on Poisson theory. These intervals are most appropriate when the response probabilities are small (usually under .10).

It should be noted that these procedures do not take the monotonic nature of the response probabilities into account. We consider each of these approaches in turn.

B. Method 1 Asymptotic Approach

To use the asymptotic normal approach we adopt a conservative yardstick and require that each cell in the 2 x 2 table under consideration contain at least 5 responses.

For most situations of practical interest both p_1 and p_4 will be relatively far from 1. Certainly if p_1 , the mortality rate in the control group is close to 1, the test will be terminated. If p_4 is very close to 1 while p_1 is close to 0, there is no need in calculating confidence intervals on their ratio. Group 4 will be obviously unsatisfactory.

We wish to calculate an asymptotic theory confidence interval on the ratio

$$\theta = p_4/p_1$$

Let

$$\phi \equiv \ln \theta = \ln p_4 - \ln p_1$$

We estimate ϕ by

$$\hat{\phi} \equiv \ln \hat{p}_4 - \ln \hat{p}_1$$

As $N_1, N_4 \rightarrow \infty$ with p_1, p_4 fixed

$$\hat{\phi} \text{ is approximately } N\left(\phi, \frac{q_1}{N_1 p_1} + \frac{q_4}{N_4 p_4}\right)$$

Thus an approximate 95% confidence interval on ϕ is

$$\hat{\phi} - 1.96 \left[\frac{q_1}{N_1 p_1} + \frac{q_4}{N_4 p_4} \right]^{1/2} \leq \phi \leq \hat{\phi} + 1.96 \left[\frac{q_1}{N_1 p_1} + \frac{q_4}{N_4 p_4} \right]^{1/2}$$

In the case of the Holcombe and Phipps example

$$\begin{aligned} N_1 = N_4 &= 100 & \hat{q}_1 &= 1 - .06 = .94 \\ \hat{p}_1 &= 6/100 = 0.06 & \hat{q}_4 &= 1 - .13 = .87 \\ \hat{p}_4 &= 13/100 = 0.13 \end{aligned}$$

Substituting $\hat{p}_1, \hat{q}_1, \hat{p}_4, \hat{q}_4$ for the corresponding parameters in the standard error formula we have

$$\left[\frac{\hat{q}_1}{N_1 \hat{p}_1} + \frac{\hat{q}_4}{N_4 \hat{p}_4} \right]^{1/2} = \left[\frac{.94}{100(.06)} + \frac{.87}{100(.13)} \right]^{1/2} = 0.473$$

Thus an approximate 95% confidence interval on ϕ is

$$(0.773 - 1.96(.473), 0.773 + 1.96(.473)) = (-.154, 1.700)$$

Therefore, $(e^{-.154}, e^{1.700})$ is an asymptotic 95% confidence interval on $\theta \equiv p_4/p_1$.

This interval is

$$(0.857, 5.474)$$

The conclusions from this confidence interval calculation are

- p_4 is not "significantly" different from p_1 at the .05 level since the confidence interval contains 1. (Note that we observed borderline significance with Williams' procedure at $\alpha = 0.05$).
- p_4 is not very much smaller than p_1 (at least 86% of p_1) but may be much larger than p_1 (as much as 5.5 times p_1)
- p_4/p_1 is not determined very precisely by the data, based on such a comparison.

We have thus quantified the relation between p_1 and p_4 .

We now calculate 95% confidence intervals to compare the response rates in each of the other treatment groups with that in the control group.

	CONTROL	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6
DEAD	6	8	8	13	79	100
LIVE	94	92	92	87	21	0
	100	100	100	100	100	100

Holcombe and Phipps Compound D Fry Mortality Data Pooled Across
Tanks within Treatment Groups

We are not particularly interested in comparing group 6 with the control since it is obviously inferior. We thus compare groups 2, 3, and 5 with the control group by means of asymptotic 95% confidence intervals.

Group 2 vs Control

$$\theta \equiv p_2/p_1$$

$$\phi \equiv \ln \theta = \ln p_2 - \ln p_1$$

$$\hat{\phi} = \ln \hat{p}_2 - \ln \hat{p}_1 = \ln 0.08 - \ln 0.06 = 0.288$$

$$\text{stderr}(\hat{\phi}) = \left[\frac{\hat{q}_2}{N_2 \hat{p}_2} + \frac{\hat{q}_1}{N_1 \hat{p}_1} \right]^{1/2} = \left[\frac{.92}{100(.08)} + \frac{.94}{100(.06)} \right]^{1/2} = 0.521$$

$$\hat{\phi} \pm 1.96 \text{ stderr}(\hat{\phi}) = 0.288 \pm 1.96(0.52) = (-.733, 1.309)$$

Thus an asymptotic 95% confidence interval on θ is $(e^{-.733}, e^{1.309}) = (0.480, 3.703)$. This implies that there is no statistical evidence at the .05 level of a difference between p_2 and p_1 . Furthermore the present data do not determine this ratio very precisely.

Group 5 vs Control

$$\hat{\phi} = \ln \hat{p}_5 - \ln \hat{p}_1 = \ln 0.79 - \ln 0.06 = 2.578$$

$$\text{stderr}(\hat{\phi}) = \left[\frac{.21}{100(.79)} + \frac{.94}{100(.06)} \right]^{1/2} = (.159)^{1/2} = 0.399$$

$$\hat{\phi} \pm 1.96 \text{ stderr}(\hat{\phi}) = (1.794, 3.362)$$

Thus an asymptotic 95% confidence interval on θ is $(e^{1.794}, e^{3.362}) = (6.011, 28.859)$. There is thus overwhelming statistical evidence that the response rate in group 5 is substantially greater than that in the control group, by at least a factor of 6. The interval however is very

wide and so we cannot determine the ratio very precisely.

We may wish to modify these intervals for simultaneity. Since we are calculating 4 confidence intervals we can adjust their levels to attain a familywise confidence level of 0.05. The simplest way to do this is by means of Bonferroni's inequality. Namely we construct each interval at individual confidence level $1 - (.05/4) = .9875$. The appropriate normal distribution factor then becomes 2.50.

$$\exp \left\{ \ln(\hat{p}_j / \hat{p}_1) \pm 2.50 \left[\hat{q}_j / N \hat{p}_j + \hat{q}_1 / N_1 \hat{p}_1 \right]^{1/2} \right\} \quad J = 2, 3, 4, 5$$

These intervals are:

Group 2 vs Control	(.363, 4.906)
Group 3 vs Control	(.363, 4.906)
Group 4 vs Control	(.664, 7.067)
Group 5 vs Control	(4.857, 35.712)

We thus conclude that there is strong statistical evidence that group 5 has at least 5 times the response rate of group 1 but there is not enough statistical evidence to distinguish the response rates at groups 2, 3, 4, from that at group 1. Furthermore the data are not sufficient to make precise inferences about the ratios of treatment group to control group response rates without putting further structure on the problem such as assuming some sort of dose response relation. We will consider this approach in subsequent sections.

C. Method 2 Exact, Small Sample Confidence Intervals

If the sample sizes are not sufficiently large to apply the asymptotic confidence interval procedure (method 1) and if response proportions are not sufficiently small to apply Poisson theory (method 3), then confidence interval comparisons between treatment groups and control group can be made by an exact, small sample procedure. This procedure is based on the non null distribution of Fisher's exact test in 2 x 2 contingency tables.

Consider a 2 x 2 contingency table to compare the response rate in a particular treatment group with that in the control group.

	Control	Group 2	Total
Dead	X_1	X_2	$X_1 + X_2$
Live	$m - X_1$	$m - X_2$	$m + n - (X_1 + X_2)$
	m	n	$m + n$

Let p_1, p_2 denote the response probabilities (e.g. probability of death) within the control group and treatment group respectively. We can test the hypothesis

$$H_0: p_1 = p_2$$

$$\text{vs } H_1: p_1 \neq p_2$$

by means of Fisher's exact test (Lehmann [25], Lieberman and Owen [26]), conditional on the margins of the table being fixed. This test is based on the hypergeometric distribution. We reject H_0 if X_2 is too extreme.

The nonnull distribution of X_2 conditional on $X_1 + X_2 = t$ is

$$P(X_2 = x | X_1 + X_2 = t) = C_t(\rho) \binom{m}{t-x} \binom{n}{x} \rho^x \quad x = 0, 1, 2, \dots, t$$

where $C_t(\rho)$ is a normalizing constant and

$$\rho = \frac{p_1/q_1}{p_2/q_2}$$

The quantity ρ is known as the odds ratio and is very important for power calculations and for calculating confidence intervals to compare the response rates in the treatment and control groups on a pairwise basis.

The odds ratio is a quantity between 0 and ∞ . $\rho = 1$ if and only if $p_1 = p_2$. If $\rho > 1$ then $p_1/p_2 > 1$ and if $\rho < 1$ then $p_1/p_2 < 1$. The size of the confidence interval on ρ indicates how precisely this quantity can be estimated from the data.

Thomas [39] presents an algorithm for calculating exact, small sample confidence intervals on ρ based on the distribution of X_2 , conditional on the margins of the table. We have implemented Thomas' algorithm in EXAX2 [14] and illustrate the calculation of the confidence intervals with several examples.

We first consider the Holcombe and Phipps compound D fry mortality data. The output appears in Figure XIII.1. The first page of the output defines the odds ratio explicitly in terms of the order of the groups and the order of the response categories. Subsequent pages present the individual 2×2 tables to compare treatment groups with the control group on a pairwise basis, a point estimate and confidence interval on the odds ratio and the one sided significance level of Fisher's exact test for equality of the two response probabilities.

It should be noted that the quantities ALPHAL and ALPHAU, which specify the probability inequalities governing the upper and lower

confidence limits are under the control of the user. They can be adjusted to yield one sided upper or lower confidence bounds in place of two sided intervals or to account for simultaneity by means of Bonferroni's method.

In the present example individual 95% two-sided confidence intervals are calculated on the odds ratios of each treatment group with the control group. The conclusions are similar to those arrived at with the asymptotic intervals. Namely the response rates in groups 2 and 3 cannot be distinguished from that in the control group. The fry mortality rate in group 4 is marginally worse than the control group rate. The lower confidence limit of 0.13 suggests that the fry mortality rate in group 4 could be substantially worse than the control rate. The upper confidence limit of 1.27 is not too far removed from 1.0. This implies that the fry mortality rate in group 4 is not significantly different from that in group 1 at $\alpha = 0.05$ but would be significant at a slightly higher α -level. ($\alpha = 0.07$ suffices here). The odds ratios comparing the response rates in groups 5, 6 to that in the control group are very small and the upper bounds are very small. There is thus strong evidence that these groups have significantly higher fry mortality rates than the control group and substantially so.

The large widths of the confidence intervals imply that the odds ratios cannot be determined very precisely.

We next consider the Holcombe and Phipps compound D embryo mortality data. The output format is the same as that for the fry mortality data and appears in Figure XIII.2. We see that none of the treatment group response rates are significantly different from the control group rate. The confidence intervals all straddle 1 and so the treatment group response rates cannot be distinguished from the control group response rate. This is in conformance with the results of our preliminary analyses.

The previous discussion pertained to construction of exact, small sample confidence intervals on the odds ratio

$$\rho = \frac{p_1/q_1}{p_2/q_2}$$

However ρ has no direct physical interpretation. A parameter such as

$$\theta = p_2/p_1$$

is more physically meaningful. How can we construct confidence intervals on θ based on the confidence intervals we have constructed on ρ ? We can express θ in terms of ρ and p_1 . Namely

$$\theta = \frac{1}{\rho + p_1(1 - \rho)}$$

If $\rho < 1$ then θ decreases as p_1 increases from 0 to 1.

If $\rho > 1$ then θ increases as p_1 increases from 0 to 1.

For fixed p_1 , θ decreases as ρ increases from 0 to ∞

Suppose $(\underline{\rho}, \bar{\rho})$ is a confidence interval on ρ and suppose

$(\underline{p}_1, \bar{p}_1)$ is a confidence interval on p_1 .

Then a conservative confidence interval on θ is

$$\left(\frac{1}{\bar{\rho} + \bar{p}_1^*(1 - \bar{\rho})}, \frac{1}{\underline{\rho} + \underline{p}_1^{**}(1 - \underline{\rho})} \right)$$

Where

$$p_1^* = \begin{cases} \bar{p}_1 & \text{as } \bar{\rho} > 1 \\ \underline{p}_1 & \text{as } \bar{\rho} < 1 \end{cases} \quad p_1^{**} = \begin{cases} \underline{p}_1 & \text{as } \underline{\rho} < 1 \\ \bar{p}_1 & \text{as } \underline{\rho} > 1 \end{cases}$$

The confidence interval on the odds ratio ρ comes from the EXAX2 program output. Confidence intervals on p_1 can be calculated by the Pearson-Clopper method. Namely if

$$\hat{p}_1 = \frac{X_1}{N_1} \quad \text{then}$$

$$\underline{p}_1 = \left\{ 1 + \frac{N_1 - X_1 + 1}{X_1} F(2N_1 - 2X_1 + 2, 2X_1; 1 - \alpha/2) \right\}^{-1}$$

$$= 0 \text{ if } X_1 = 0$$

$$\bar{p}_1 = \left\{ 1 + \frac{N_1 - X_1}{X_1 + 1} \frac{1}{F(2X_1 + 2, 2N_1 - 2X_1; 1 - \alpha/2)} \right\}^{-1}$$

$$= 1 \text{ if } X_1 = N_1$$

These confidence intervals are given in chart form. See for example Box, Hunter, and Hunter [40], pages 642, 643 or Dixon and Massey [13], pages 501-504.

We apply this conservative procedure to the Holcombe and Phipps compound D fry mortality data and compare the results with those calculated by the asymptotic approach.

In the control group $X_1 = 6$, $N_1 = 100$.

Thus $\hat{p}_1 = 0.06$. A 99% 2 sided confidence interval on p_1 is, from the Pearson-Clopper charts entered at $\hat{p} = 0.06$, $n = 100$, $(0.02, 0.15) \equiv (\hat{p}_1, \tilde{p}_1)$.

The 95% confidence intervals on the odds ratio ρ , namely $(\hat{\rho}, \tilde{\rho})$ are

Group 2 vs Control (0.2018, 2.5238)

Group 3 vs Control (0.2018, 2.5238)

Group 4 vs Control (0.1278, 1.2743)

Group 5 vs Control (0.0055, 0.0467)

Group 6 vs Control (0, 0.0027)

Combining these results as discussed previously, we obtain:

Groups 2, 3 vs Control

Since $\tilde{\rho} > 1$, $\hat{\rho} < 1$ we have

$$\hat{\theta} = \frac{1}{\tilde{\rho} + \hat{p}_1(1 - \tilde{\rho})} = \frac{1}{2.5238 + 0.02(1 - 2.5238)} = 0.401$$

$$\tilde{\theta} = \frac{1}{\hat{\rho} + \hat{p}_1(1 - \hat{\rho})} = \frac{1}{0.2018 + 0.02(1 - 0.2018)} = 4.592$$

This would be a conservative $100(1 - .05 - .01) = 94\%$ confidence interval.

The corresponding 95% confidence interval based on asymptotic normal theory (0.480, 3.703). We see that the two intervals are qualitatively similar but that the conservative interval is longer, as would be expected.

We now compare the conservative small sample with the approximate large sample intervals for comparisons of groups 4, 5, 6 with the control group. The calculations proceed analogously.

	<u>Conservative, Small Sample</u> <u>Interval on θ</u>	<u>Approximate, Large Sample</u> <u>Interval on θ</u>
Groups 2,3 vs Control	(0.401, 4.592)	(0.480, 3.703)
Group 4 vs Control	(0.788, 6.885)	(0.857, 5.474)
Group 5 vs Control	(5.272, 39.856)	(6.011, 28.859)
Group 6 vs Control	(6.566, 50)	large sample interval not calculated

We see that the two sets of intervals are qualitatively similar however the conservative, small sample intervals are 30%-51% longer than the corresponding asymptotic intervals.

An alternative approximation can be used to calculate conservative confidence intervals on $\theta \equiv p_2/p_1$. Consider again the 2 x 2 table.

	CONTROL	GROUP 2
DEAD	X_1	X_2
LIVE	Y_1	Y_2
	N_1	N_2

Let p_1, p_2 denote the probabilities of death in groups 1, 2 respectively. We wish to construct a confidence interval on $p_2/p_1 \equiv \theta$.

Now N_1, N_2 were fixed by the experimenter. Let $r \equiv N_2/N_1$. Suppose we assume the fiction that $N_1 \sim P_0(\lambda)$, $N_2 \sim P_0(r\lambda)$ and that N_1, N_2 in the data are realizations of these two independent random variables. Then X_1, X_2, Y_1, Y_2 can be treated as independent Poisson random variables with means $p_1\lambda, p_2\lambda, q_1\lambda, q_2r\lambda$ respectively. Confidence intervals on p_2/p_1 can be constructed by methods like those discussed in connection with the Poisson approximation approach, (method 3). Namely

$$P \frac{X_2}{X_1 + 1} \frac{1}{F(2X_1 + 2, 2X_2; 1 - \alpha_1)} \frac{1}{r} \leq \frac{p_2}{p_1} \leq \frac{X_2 + 1}{X_1} F(2X_2 + 2, 2X_1; 1 - \alpha_2) \frac{1}{r} \geq 1 - \alpha$$

where $\alpha_1 + \alpha_2 = \alpha$. Now these confidence intervals are conservative because we are introducing additional variability by assuming that N_1, N_2 are random variables rather than fixed constants. The variances of X_1, X_2

are inflated from $N_1 p_1 q_1$, $N_2 p_2 q_2$ to $N_1 p_1$, $N_2 p_2$ by this assumption. Thus the greater are p_1 , p_2 , the more conservative this procedure will be.

We illustrate the application of these intervals with the Holcombe and Phipps compound D fry mortality data and the DeFoe 1, 1, 2, trichloroethane fry mortality data.

First consider the Holcombe and Phipps compound D fry mortality data. The comparisons of Groups 2, 3, 4 vs Control, based on the Poisson approximation, are quite similar to the conservative small sample confidence intervals discussed earlier in this subsection.

Now consider comparisons of Groups 5, 6 with the Control group.

Group 5 vs Control: $X_5 = 79$, $X_1 = 6$, $N_5 = N_1 = 100$, $\alpha_1 = \alpha_2 = 0.025$

Thus

$$\left(\frac{79}{7} \frac{1}{F(14, 158; .975)}, \frac{80}{6} F(160, 12; .975) \right) = \left(\frac{79}{7} (1.94)^{-1}, \frac{80}{6} (2.77) \right) = (5.82, 36.93)$$

is an approximate 95% confidence interval on p_5/p_1 .

Group 6 vs Control: $X_6 = 100$, $X_1 = 6$, $N_6 = N_1 = 100$, $\alpha_1 = \alpha_2 = 0.025$

Thus

$$\left(\frac{100}{7} \frac{1}{F(14, 200; .975)}, \frac{101}{6} F(202, 12; .975) \right) = \left(\frac{100}{7} \frac{1}{1.79}, \frac{101}{6} (2.75) \right) = (7.98, 46.29)$$

is an approximate 95% confidence interval on p_6/p_1 .

These intervals compare with the conservative, small sample intervals calculated earlier as follows:

	<u>Conservative, Small Sample</u>	<u>Approximate Poisson</u>
Group 5 vs Control	(5.27, 39.86)	(5.82, 36.93)
Group 6 vs Control	(6.57, 50)	(7.98, 46.29)

These intervals are seen to be quite similar.

We now consider the DeFoe 1, 1, 2, trichloroethane data and calculate approximate confidence intervals to compare Groups 5, 6 to the control group. Since $X_1 = 0$ we can only calculate lower confidence bounds.

Group 5 vs Control: $X_5 = 9$, $X_1 = 0$, $N_5 = N_1 = 40$, $\alpha_1 = 0.05$, $\alpha_2 = 0$.

Thus

$$\frac{9}{1} \frac{1}{F(2, 20; .95)} = \frac{9}{3.49} = 2.58$$

is a 95% lower confidence bound on p_5/p_1 .

Group 6 vs Control $X_6 = 40$, $X_1 = 0$, $N_6 = N_1 = 40$, $\alpha_1 = 0.05$, $\alpha_2 = 0$.

Thus

$$\frac{40}{1} \frac{1}{F(2, 82; .95)} = \frac{40}{3.12} = 12.82$$

is a 95% lower confidence bound on p_6/p_1 .

Thus there is strong statistical evidence that the response rates in groups 5 and 6 are substantially greater than that in the control group. The response rate in group 5 is at least $2\frac{1}{2}$ times that in the control group.

D. Method 3 Poisson Approximation

We now consider method 3 for placing confidence intervals on ratios of parameters. This method is based on the Poisson approximation to the binomial distribution and so requires that each p be less than 0.1 or that each p be greater than 0.9 in order that the Poisson approximation be reasonably accurate. Operationally, we will use this approximation if each \hat{p} is less than 0.1 or if each \hat{p} is greater than 0.9. The prototype situation is

	Control	Group 2
Dead	X_1	X_2
Live	$N_1 - X_1$	$N_2 - X_2$
	N_1	N_2

Let p_1 , p_2 denote the response probabilities in groups 1, 2 respectively. We wish to construct $1 - \alpha$ confidence intervals on p_2/p_1 .

Let $\lambda_1 = N_1 p_1$, $\lambda_2 = N_2 p_2$.

Then if $p_1 < .1$, $p_2 < .1$ $X_1 \sim P_0(\lambda_1)$, $X_2 \sim P_0(\lambda_2)$. We can thus pose the problem as one of placing confidence intervals on the ratio of two Poisson means. If $p_1 > .9$, $p_2 > .9$ it is probably of more interest to place a confidence interval on the ratio q_2/q_1 , where $q_1 \equiv 1 - p_1$, $q_2 \equiv 1 - p_2$. We are then back in the above situation.

Nelson [24] shows that a $1 - \alpha$ confidence interval on λ_2/λ_1 is

$$\left[\frac{X_2}{X_1 + 1} \frac{1}{F(2X_1 + 2, 2X_2; 1 - \alpha_1)}, \frac{X_2 + 1}{X_1} F(2X_2 + 2, 2X_1; 1 - \alpha_2) \right]$$

where $F(v_1, v_2; \gamma)$ represents the upper γ point of the F-distribution with d.f. v_1, v_2 and $\alpha_1 + \alpha_2 = \alpha$. Now

$$\lambda_2/\lambda_1 \equiv (N_2 p_2)/(N_1 p_1) = (N_2/N_1)(p_2/p_1)$$

Thus multiplying the above confidence bounds by the factor N_1/N_2 yields confidence bounds on p_2/p_1 . Namely

$$\left[\frac{X_2}{X_1 + 1} \frac{1}{F(2X_1 + 2, 2X_2; 1 - \alpha_1)} \frac{N_1}{N_2}, \frac{X_2 + 1}{X_1} F(2X_2 + 2, 2X_1; 1 - \alpha_2) \frac{N_1}{N_2} \right]$$

is a $1 - \alpha$ confidence interval on p_2/p_1 . Often we take α_1, α_2 to be $\alpha/2$. However for one sided confidence intervals we take $\alpha_1 = \alpha, \alpha_2 = 0$ or $\alpha_1 = 0, \alpha_2 = \alpha$.

If $X_1 = 0$ or if $X_2 = 0$ we have only one sided information about p_1, p_2 respectively. Thus we can only construct one sided confidence bounds on their ratio. Namely if $X_1 > 0, X_2 = 0$ then set the lower confidence bound equal to 0 and upper confidence bound on p_2/p_1 becomes

$$\frac{N_1}{N_2} \frac{1}{X_1} F(2, 2X_1; 1 - \alpha) \quad \text{if } X_2 = 0, X_1 > 0.$$

If $X_1 = 0, X_2 > 0$, then we can only get a lower bound on p_2/p_1 . Set the upper bound equal to ∞ and the lower confidence bound becomes

$$X_2 \frac{1}{F(2, 2X_2; 1 - \alpha)} \frac{N_1}{N_2} \quad \text{if } X_1 = 0, X_2 > 0$$

If $X_1 = 0, X_2 = 0$ the problem is indeterminate.

Nelson [24] presents charts which facilitate the construction of two sided 90%, 95% or 99% confidence intervals on λ_2/λ_1 . However his charts do not apply for the situation when $X_1 = 0$ or $X_2 = 0$. In fact they effectively apply only when $0.1 < X_2/X_1 < 10$. The charts are shown in in Figures XIII.3, XIII.4, XIII.5.

To use the Nelson charts

1. Enter the value of X_2/X_1 on the horizontal axis.
2. Go up to the curve labelled with the X_1 value. (There are two sets of curves, corresponding to upper and lower confidence limits).
3. Read the upper and lower limits on the vertical scale.
4. Multiply the resulting limits by the ratio N_1/N_2 .

We illustrate the use of this Poisson based procedure on several sets of data. First we consider the Holcombe and Phipps compound D fry mortality data. We pool responses across tanks within groups.

	CONTROL	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6
DEAD	6	8	8	13	79	100
LIVE	94	92	92	87	21	0
	100	100	100	100	100	100

We compare various treatment groups with the control group. We will calculate two sided 95 percent, nonsimultaneous confidence intervals. Groups 2 and 3 appear to have response probabilities around 0.10 and group 4 does not seem to be too much beyond this level. We thus stretch our criterion a bit and calculate confidence intervals to compare groups 2, 3, 4 with the control groups.

Note that we could modify the confidence intervals for simultaneity by using Bonferroni's inequality.

Groups 2, 3 vs Control:

$$X_2 = 8, X_1 = 6, N_2 = N_1 = 100, \alpha_1 = \alpha_2 = .025$$

Thus

$$\left(\frac{8}{7} \frac{1}{F(14, 16; .975)}, \frac{9}{6} F(18, 12; .975) \right) \equiv \frac{8}{7}(2.83)^{-1}, \frac{9}{6}(3.11) = (0.404, 4.665).$$

Group 4 vs Control:

$$X_4 = 13, X_1 = 6, N_2 = N_1 = 100, \alpha_1 = \alpha_2 = .025$$

Thus

$$\left(\frac{13}{7} \frac{1}{F(14, 26; .975)}, \frac{14}{6} F(28, 12; .975) \right) = \frac{13}{7}(2.45)^{-1}, \frac{14}{6}(2.98) = (0.76, 6.95)$$

Comparing the confidence intervals obtained by methods 1, 2, 3 we see that

	<u>Asymptotic</u>	<u>Conservative Small Sample</u>	<u>Poisson Approximation</u>
Group 2, 3 vs Control	(.480, 3.703)	(.401, 4.592)	(.404, 4.665)
Group 4 vs Control	(.857, 5.474)	(.788, 6.885)	(.76, 6.95)

Thus the asymptotic intervals are shorter than either of the small sample intervals. The small sample intervals are thus more conservative.

We next consider the DeFoe compound C fry mortality data. We again pool across tanks within groups.

	CONTROL	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6
DEAD	0	0	2	1	9	40
LIVE	40	40	38	40	31	0
	40	40	40	41	40	40

Since there are zero responses in the control group (i.e. $X_1 = 0$), we can only calculate lower confidence bounds.

Group 2 vs Control: Since Group 2 has 0 responses also, the situation is indeterminate.

Group 3 vs Control: Choose $\alpha_1 = .05$, $\alpha_2 = 0$ $X_3 = 2$, $X_1 = 0$, $N_3 = N_1 = 40$

Then

$$\frac{2}{1} \frac{1}{F(2, 4; .95)} = \frac{2}{6.94} = 0.29$$

is a 95 percent lower confidence bound on p_3/p_1 . Thus there is no statistical evidence, at the $\alpha = .05$ level, that $p_3 > p_1$.

Group 4 vs Control: $X_4 = 1$, $X_1 = 0$ $N_4 = 41$, $N_1 = 40$ Choose $\alpha_1 = .05$, $\alpha_2 = 0$.

Then

$$\frac{1}{1} \frac{1}{F(2, 2; .95)} \frac{40}{41} = \frac{40}{41} \frac{1}{19.0} = 0.05$$

is a 95 percent confidence bound on p_4/p_1 . Thus there is no statistical evidence, at the $\alpha = 0.05$ level, that $p_4 > p_1$.

In general the confidence intervals that we have calculated are too wide to determine the ratios of the various probabilities with much precision. We must conclude that the data are not sufficient to estimate these ratios very precisely without placing further structure on the problem. One way of imposing such further structure will be discussed in the following sections.

EXAX2 OUTPUT HOLCOMBE AND PHIMS COMPOUND D FRY MORTALITY DATA

EXACT 95.00% 2x2 NCI-SIMULTANEOUS CONFIDENCE INTERVALS ON ODDS RATIO OF TPT 1 VS EACH OF THE OTHER TREATMENTS.

THE ALGORITHM USED IS A MODIFIED VERSION OF AS 34 GIVEN BY THOMAS (1971) APPLIED STAT. 26:105-110. FOR EACH 2 x 2 TABLE ESTIMATES CAN BE INTERPRETED BY REFERRING TO THE TABLE BELOW.

T1	J	ALIVE		DEAD		TOTAL
		A	C	B	D	
T2	J					N1
						N2
TOTAL		M1	M2			NP

$$PSI = (P102)/(P201)$$

$$ESTIMATE OF PSI = BC/AD$$

$$ESTIMATE OF P1 = B/N1$$

$$ESTIMATE OF P2 = D/N2$$

$$ESTIMATE OF Q1 = A/N1$$

$$ESTIMATE OF Q2 = C/N2$$

Figure XIII.1 EXAX2 output from calculation of exact, small sample confidence intervals on the odds ratios between treatments and control

EXACT OUTPUT NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT

ALPHA = 0.025 ALPHA = 0.025

TPT	1	2	ALIVE	DEAD	TOTAL
			54.000	4.000	58.000
			92.000	5.000	97.000

TOTAL ALIVE = 146.000 TOTAL DEAD = 14.000 GRAND TOTAL = 160.000

ESTIMATE OF (PIQ1)/(P2Q1) = 0.734042525

95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.201835632 , 2.52381802)

PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.19140

specification of lower and upper probability levels for corresponding confidence limit calculations
 exact table compares group 1 with the control group

point estimate of odds ratio
 confidence interval on odds ratio
 observed significance level of Fisher's exact test

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT

ALPHA = 0.025 ALPHA = 0.025

TPT	1	2	ALIVE	DEAD	TOTAL
			54.000	4.000	58.000
			92.000	5.000	97.000

TOTAL ALIVE = 146.000 TOTAL DEAD = 14.000 GRAND TOTAL = 160.000

ESTIMATE OF (PIQ1)/(P2Q1) = 0.734042525

95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.201835632 , 2.52381802)

PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.19140

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT

ALPHA = 0.025 ALPHA = 0.025

TPT	1	2	ALIVE	DEAD	TOTAL
			54.000	4.000	58.000
			92.000	5.000	97.000

TOTAL ALIVE = 146.000 TOTAL DEAD = 14.000 GRAND TOTAL = 160.000

ESTIMATE OF (PIQ1)/(P2Q1) = 0.734042525

95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.201835632 , 2.52381802)

PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.19140

Figure XIII.1 Continued

EXACT OUTPUT NONCONC AND PHARS COMPOUND O FRY MORTALITY DATA

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 5

ALPHA = 0.025 ALPHAU = 0.025

TRT	1	5	ALIVE	DEAD	TOTAL
TRT 1	1	5	94.000	6.000	100.000
TRT 5	1	5	21.000	79.000	100.000

TOTAL ALIVE = 115.000 , TOTAL DEAD = 85.000 , GRAND TOTAL = 200.000

ESTIMATE OF (P1231/122011) = 0.0149474284
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.0054938793, 0.0467329025)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.000000

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 6

***** ODDS RATIO ESTIMATE IS ZERO, ONLY AN UPPER CONFIDENCE BOUND CAN BE OBTAINED. *****

ALPHA = 0.0 ALPHAU = 0.050

TRT	1	6	ALIVE	DEAD	TOTAL
TRT 1	1	6	94.000	6.000	100.000
TRT 6	1	6	0.0	100.000	100.000

TOTAL ALIVE = 94.000 , TOTAL DEAD = 106.000 , GRAND TOTAL = 200.000

ESTIMATE OF (P1231/122011) = 0.0
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.0, 0.0026908517)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.000000

Figure XIII.1 Continued

EXAX2 output Halconge and Phins compound D embryo mortality data

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 2

ALPHA = 0.025 ALPHA0 = 0.025

TRT	1	2	ALIVE	DEAD	TOTAL
TRT 1	130.000	70.000	200.000		
TRT 2	130.000	70.000	200.000		
TOTAL	ALIVE = 260.000	TOTAL = 140.000	GRAND TOTAL = 400.000		

ESTIMATE OF (P100)/(P200) = 1.0490040
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.64900069 , 1.54032756)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.54174

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 3

ALPHA = 0.025 ALPHA0 = 0.025

TRT	1	3	ALIVE	DEAD	TOTAL
TRT 1	130.000	70.000	200.000		
TRT 3	137.000	63.000	200.000		
TOTAL	ALIVE = 267.000	TOTAL = 133.000	GRAND TOTAL = 400.000		

ESTIMATE OF (P100)/(P200) = 1.17000945
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.75745113 , 1.81522560)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.26219

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TRT 1 VS TRT 4

ALPHA = 0.025 ALPHA0 = 0.025

TRT	1	4	ALIVE	DEAD	TOTAL
TRT 1	130.000	70.000	200.000		
TRT 4	122.000	78.000	200.000		
TOTAL	ALIVE = 252.000	TOTAL = 148.000	GRAND TOTAL = 400.000		

ESTIMATE OF (P100)/(P200) = 0.84200041
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = (0.54000330 , 1.29054032)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.23429

Figure XIII.2 EXAX2 output from calculation of exact, small sample confidence intervals on treatment-control odds ratio

EXTRA OUTPUT WOLCUMBE AND PHANS COMPARED D EMBRYO MORTALITY DATA

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT 6

ALPHA = 0.025 ALPHA0 = 0.025

TPT	1	ALIVE	DEAD	TOTAL
TPT	6	130.000	70.000	200.000
		133.000	67.000	200.000
TOTAL	ALIVE	263.000	TOTAL	DEAD
			137.000	GRAND TOTAL
				400.000

ESTIMATE OF $(P102)/(P201) = 1.0688540$
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = 1 0.692345023 , 1.65059376)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.41639

EXACT 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL ON ODDS RATIO OF TPT 1 VS TPT 6

ALPHA = 0.025 ALPHA0 = 0.025

TPT	1	ALIVE	DEAD	TOTAL
TPT	6	130.000	70.000	200.000
		124.000	76.000	200.000
TOTAL	ALIVE	254.000	TOTAL	DEAD
			146.000	GRAND TOTAL
				400.000

ESTIMATE OF $(P102)/(P201) = 0.878542493$
 95.00 % NON-SIMULTANEOUS CONFIDENCE INTERVAL = 1 0.572517395 , 1.34771442)
 PROBABILITY OF TABLE UNDER HYPOTHESIS OF INDEPENDENCE = 0.30182

Figure XIII.2 Continued

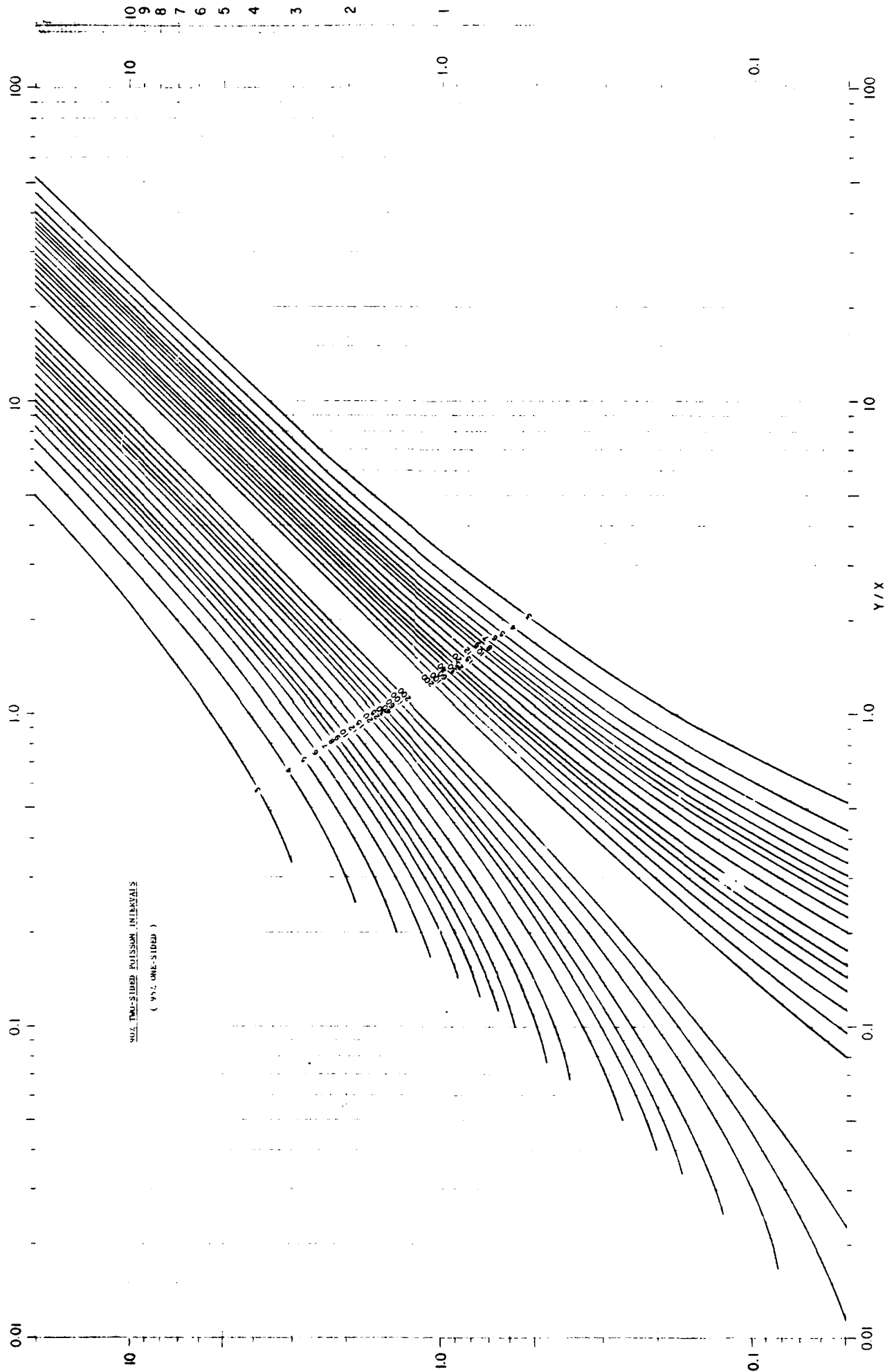


Figure XIII.3 Poisson two sample confidence interval charts -- 90 percent (from Nelson, 1969)

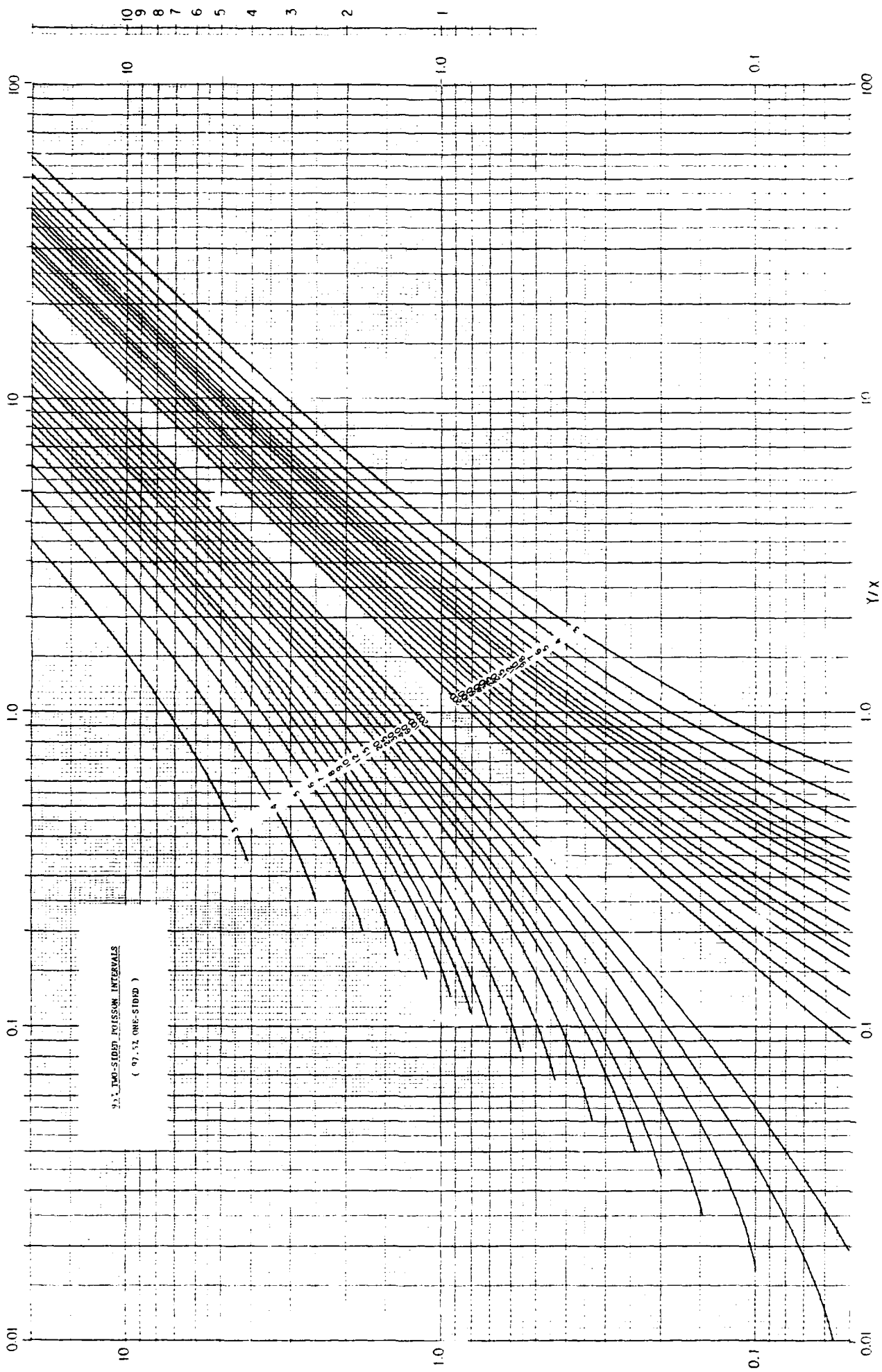


Figure XIII.4 Poisson two sample confidence interval charts -- 95 percent (from Nelson, 1969)

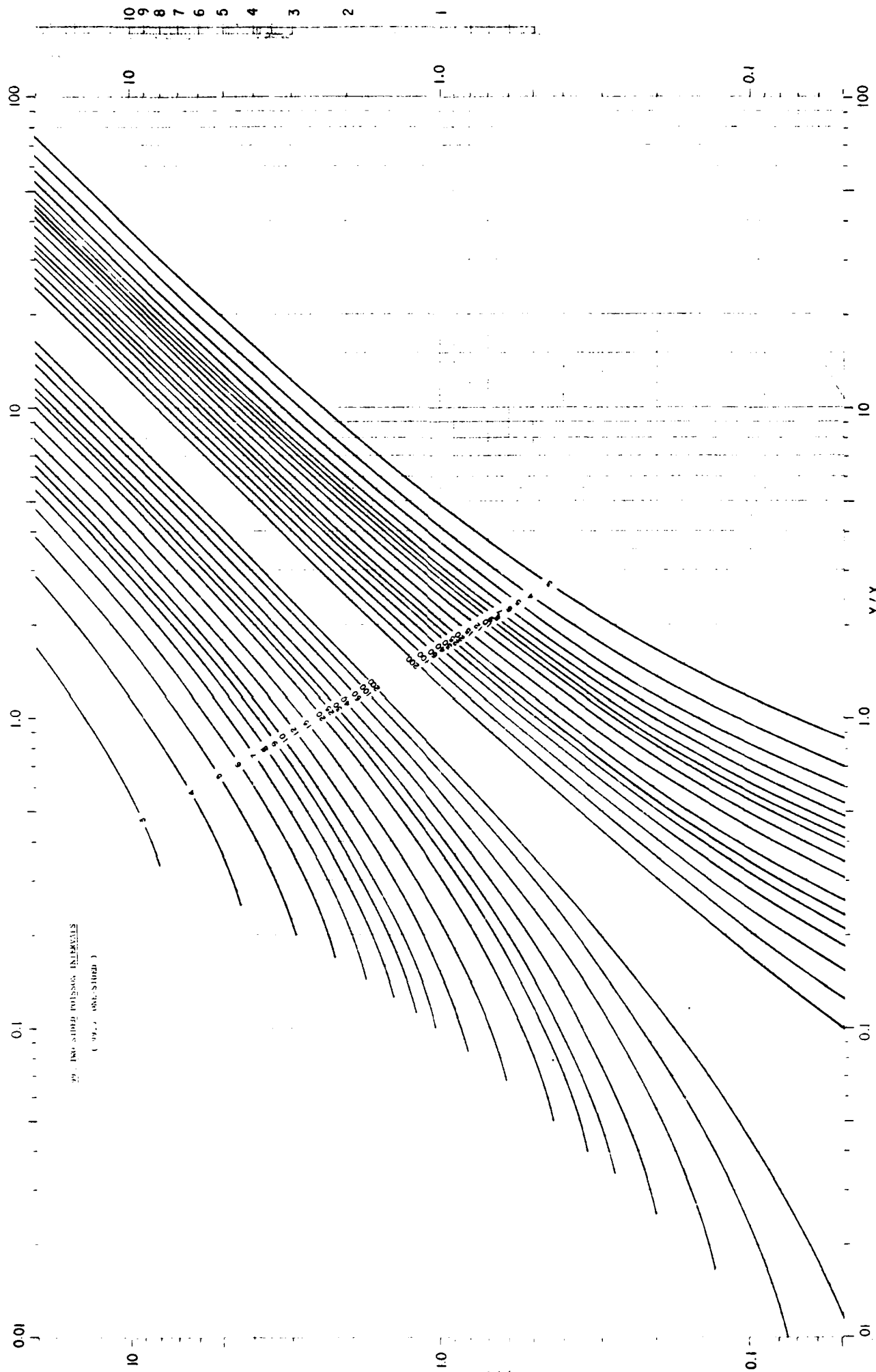


Figure XIII.5 Poisson two sample confidence interval charts -- 99 percent (from Nelson, 1969)

XIV. DOSE RESPONSE CURVE ESTIMATION -- PROBIT ANALYSIS

A. Introduction. Dose Response Estimation vs Hypothesis Testing

An alternative approach to estimating acceptable concentration levels is based on fitting dose response models to the data and estimating that concentration, C_L , which results in an increase of at most L in the response rate over and above background level. The dose response curve formulation is pictured schematically in Figure XIV.1. The problem of determining a safe concentration has been transformed from a testing problem (determine which response rates are significantly different than the control rate) to an estimation problem (calculate a lower confidence bound on C_L).

The two formulations are conceptually different and lead to different implications. With the classical hypothesis testing formulation the larger and more precise the experiment the more powerful will be the hypothesis test. Thus lower concentration levels will be found significantly different from the control group and so the acceptable concentration will be decreased. By contrast, with the dose response curve estimation formulation the larger and more precise the experiment, the higher will be the lower confidence bound on C_L and so the acceptable concentration will be increased. This latter situation seems more natural to us for two reasons.

1. There is no need to specify rigid sample size requirements in the protocol. People could present any level of evidence regarding safe concentrations that they wish. The more extensive the experiment, the higher will be the lower confidence bound on C_L .
2. An investigator conducting toxicity tests in support of petitions to the EPA for discharge permits is induced to carry out more extensive and more precise experimentation by the economics of the situation. He is rewarded for his efforts by demonstrating a greater safe concentration.

OPINION: We feel that increased emphasis should be placed on the fitting and use of dose response curve models in the design of and analysis of data from aquatic toxicity tests.

It should be noted that just because we define C_L in terms of the concentration associated with an increase in response rate of L units over background does not mean that we consider killing 100L percent of the fish to be "acceptable". No increased mortality is really desirable. However by adopting this formulation we can argue that we are limiting our risk to an upper bound on L. The choice of L in a particular situation would of course need to be a biological and a regulatory decision.

We have fitted (or attempted to fit) a number of dose response models to the embryo and fry mortality data. Some of these models are standard while others are nonstandard. Among the standard models fitted are the probit model (Finney [11]) with either logarithmic or untransformed concentration and the logit model with either logarithmic or untransformed concentration levels. Both of these models classically account for background variation by means of Abbott's correction. For example a probit model with Abbott's correction might state

$$p(\text{conc}) = p_0 + (1 - p_0) \Phi (\beta_0 + \beta_1 \ln (\text{conc}))$$

where p_0 , $p(\text{conc})$ are the response rates at the control and at conc respectively, $\Phi(\cdot)$ is the normal c.d.f., and p_0 , β_0 , β_1 are unknown parameters to be estimated from the model fit. Such a probit model can easily be fitted to the data using SAS PROC PROBIT [12]. The 1979 version of the BMDP package [27] contains a stepwise logistic regression program.

Among nonstandard dose response models tried are a nonstandard probit type model and a "nonparametric" dose response model. The nonstandard probit type model differs from the standard model in the way it handles background response. One version can be written as

$$p(\text{conc}) = \Phi (\alpha_0 + \alpha_1 \ln (\text{conc} + c))$$

where $p(\text{conc})$ is the response rate at conc, c accounts for the background response, and α_0 , α_1 , c are unknown parameters to be estimated from the model fit. A criticism of Abbott's correction is that it tacitly assumes that background related response and toxicant related response are due to different and independent mechanisms. The nonstandard model assumes that background related responses and toxicant related responses are due to similar mechanisms and thus that background acts like an incremental toxicant level c . Which (if either) model is more appropriate in a given situation depends on how well they fit the data and on biological judgement. The nonstandard probit model and a large family of other standard and nonstandard dose response models can be fitted by the use of nonlinear regression programs such as SAS PROC NLIN [12] and BMDP programs BMDP3R, BMDPAR [27] (program versions 1977 or later).

We have developed a "nonparametric" dose response model that differs from the more usual parametric models in a number of ways.

1. There is no need to make strong parametric assumptions about the form of the dose response model.
2. There is no need to be concerned with transformations of the concentration levels.
3. There is no need to worry about the parametric form used to correct for background variation.

4. Exact, small sample theory is used to construct conservative lower bounds on safe concentration.

We have developed a special purpose computer program to carry out such nonparametric dose response analyses. It is described in detail in Feder and Sherrill [41], which is included as an appendix to section XVI.

We now illustrate inferences about safe concentrations based on the various dose response models discussed above.

B. Probit Analysis Using SAS PROC PROBIT

In this subsection we fit probit models to the fry mortality data from the DeFoe test on compound C and from the Holcombe and Phipps test on compound D. Although we do not adjust the data for tank to tank heterogeneity, the same analyses can be carried out after such adjustments have been made.

We first consider the DeFoe data. The basic data are listed in Figure XIV.2. There are two tanks per treatment group. Concentration values for each group (in units of $\mu\text{g/liter}$) have been determined as average values over all determinations and over all tanks within each group. These are denoted as CONCMEAN. Other variables of importance are

DEADESUM \equiv # dead embryos in the tanks after hatch. (After about 5 days).

DEADFSUM \equiv # dead fry after 32 days.

PRPDEADE, PRPDEADF \equiv proportions of dead embryos and fry respectively.

LOGCONC $\equiv \log_{10}(\text{CONC})$

Note that the measured concentration in the control group is not zero here and that no fry mortality has occurred in the control group. It is unclear from preliminary plots of proportions of dead fry vs arithmetic and logarithmic concentration (not shown) whether a probit model would better be fitted to arithmetic or to logarithmic concentration. We will try both fits and compare them.

We first fit a standard probit model using arithmetic concentration. The specific model fitted is

$$p(\text{CONC}) = c + (1 - c) \Phi(\beta_0 - 5 + \beta_1 \text{CONC})$$

where $\Phi(\cdot)$ is the standard normal c.d.f., β_0 and β_1 are unknown

model parameters that characterize the shape of the response curve, and c is the unknown model parameter that specifies the background rate. (The quantity 5 in the argument of $\Phi(\cdot)$ is due to probit convention). We fit this model to the data with SAS PROC PROBIT by maximum likelihood estimation. The output resulting from this fit appears in Figures XIV.3 - XIV.6. The interpretation of this output is as follows:

- ① Summary of the maximum likelihood iteration process.

Intercept, Slope $\leftrightarrow \beta_0, \beta_1$ respectively in the probit model.

$c \leftrightarrow$ background rate, or threshold rate.

(Note that if c goes negative at an iteration step it is set to 0.).

MU, SIGMA (μ, σ) correspond to the mean and standard deviation of the dose-response distribution.

μ, σ are related to β_0, β_1 as

$$\beta_0 = 5 - \mu/\sigma, \beta_1 = 1/\sigma$$

These relations can be verified from the entries given in the output.

- ② The estimated asymptotic variance-covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1, \hat{c})$. This is based on the Fisher information's inverse.
- ③ The estimated asymptotic variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{c})$. This is based on the inverse of the estimated Fisher information matrix.

Note: These estimated var-cov matrices are the basis of the confidence interval calculations made by the program. The validity of these var-cov estimates depends on having the true state of nature and the maximum likelihood estimates interior to the parameter space. In this fit $\hat{c} = 0.003$ with an estimated standard error of 0.03. We thus might consider dropping c from the model.

- ④ Chi square test for lack of fit of the probit model.

Degrees of freedom \equiv number of groups - number of parameters $\equiv 6 - 3 = 3$.

Under the null hypothesis of no lack of fit to the model this statistic has a chi square distribution with 3 d.f.

- ⑤ A plot of the fitted straight line in the probit domain, with the estimated probits of the dose response rates at each concentration level in the data indicated as X's on the plot.

Notes: 1. Probit (0) $\equiv 5 + \Phi^{-1}(0) = -\infty$. However it is plotted as 0 because that is the smallest value used. Similarly, Probit (1) $\equiv 5 + \Phi^{-1}(1) = \infty$, but it is plotted as 10 because that is the largest value used.

2. The observed thresholds at probit values of 0 and 10 seem far away from the fitted line. The standard errors of these points are also very large, so these points are discounted when determining a probit fit. In particular,

$$\text{Var}[\text{probit}(\hat{p})] = p(1-p)/n\phi(\Phi^{-1}(p)) \rightarrow \infty \text{ as } p \rightarrow 0 \text{ or } 1.$$

Thus these points carry very little weight in the straight line fit in the probit domain.

3. The estimated background response rate has been removed from the plot. Thus estimates represent increments over background.

- ⑥ Plot of $\Phi(\hat{\beta}_0 - 5 + \hat{\beta}_1 \text{ CONC})$ vs CONC.

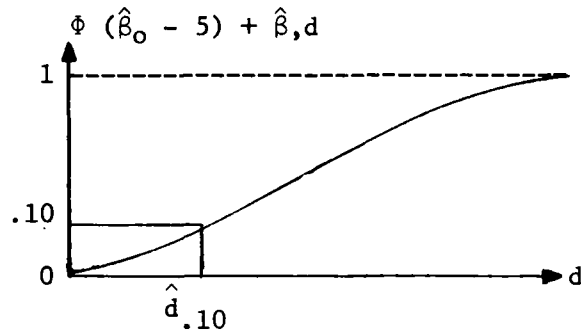
- ⑦ For various percentiles of the fitted dose response curve (after adjusting for background), the point estimates of CONCMEAN are given as well as 95% lower and upper confidence bounds on these points.

Note: These percentiles are percentages of the population responding due to the toxicant, after adjusting for background effects.

The point estimates correspond to the percentiles indicated on the plots.

These lower confidence bounds are just the quantities needed to calculate confidence bounds on safe concentrations. Lower 95% two sided bounds correspond to lower 97.5% one sided bounds. Suppose we are willing to tolerate an increase in response of 10 percent due to toxicant causes. What is a lower confidence bound on safe dose?

Consider the dose response curve (adjusted for background rate).



Now $\hat{d}_{.10} = \frac{\Phi^{-1}(.10) + 5 - \hat{\beta}_0}{\hat{\beta}_1} = 32.65$ from Figure XIV.6.

The lower 97.5 percent confidence interval on $d_{.10}$ is -45.97 . It is thus totally uninformative due to the gentle slope of the dose response curve (0.036) and relatively large standard error of the slope (0.014).

Note that the confidence bounds on the percentiles of the dose response curve are based on Fieller's theorem. See Finney [11], section 4.7 (esp. eqns (4.37), (4.38)) for details.

We now consider the chi square test statistic for goodness of fit in more detail. The chi square statistic can be used for a number of purposes. The statistic given in (4) is

$$\text{CHI-SQUARE} = \sum_{i=1}^6 \frac{(O_i - N_i \hat{p}_i)^2}{N_i \hat{p}_i \hat{q}_i}$$

Where O_i = # observed responses in the i -th treatment group

N_i = # fish in the i -th treatment group

\hat{p}_i = estimated response probability in the i -th treatment group.

$\hat{q}_i = 1 - \hat{p}_i$.

$$\hat{p}_1 = \hat{c} + (1 - \hat{c}) \Phi \left((\hat{\beta}_0 - 5) + \hat{\beta}_1 d_1 \right) \equiv$$

$$\hat{c} + (1 - \hat{c}) \Phi \left(\frac{d_1 - \hat{\mu}}{\hat{\sigma}} \right) \quad i = 1, \dots, 6.$$

The values of these quantities for the 6 treatment groups are as follows:

Trt Grp	<u>0</u>	<u>N</u>	<u>\hat{p}</u>	<u>$N\hat{p}$</u>	<u>$N\hat{p}\hat{Q}$</u>	$\frac{(0 - N\hat{p})^2}{N\hat{p}\hat{Q}}$
1	0	40	0.0100	0.3996	0.3956	0.4036
2	0	40	0.0115	0.4586	0.4534	0.4639
3	2	40	0.0154	0.6152	0.6058	3.1655
4	1	41	0.0300	1.2300	1.1931	0.0443
5	9	40	0.2397	9.5876	7.2896	0.0474
6	40	40	0.9979	39.9144	0.0854	0.0858
						$\chi^2 \equiv 4.2105$

We see that this chi square statistic agrees with that calculated in (4) of the PROBIT output. We should break out the cell by cell contributions in order to ensure that a large value of chi square is not due to one or a few cells with very low expected frequency. Just one observed response in such a cell can inflate the chi square statistic tremendously. In our case this does not occur.

Note that the applicability of the asymptotic chi square approximation to the distribution of χ^2 is doubtful here due to the small expected sample sizes. Namely

i	1	2	3	4	5	6
$N\hat{p}_i$	0.40	0.46	0.62	1.23	9.59	39.91
$N\hat{q}_i$	39.60	39.54	39.38	39.77	30.41	0.09

Dixon and Massey [13] page 238 state that for the approximate asymptotic χ^2 distribution to be close "the sample size N must be sufficiently large that none of the F_i 's (i.e. $N_i\hat{p}_i$ or $N_i\hat{q}_i$) is less than 1 and not more than 20 per cent of the F_i 's are less than 5." This criterion is clearly not met in the above example.

Since no control group mortality was observed and since the estimated background rate is compatible with 0 ($\hat{c} = 0.0031$, $\text{stderr}(\hat{c}) = .0288$) it was decided to refit the model specifying that $c = 0$. This simplification will reduce the standard errors of estimates considerably.

The fit is shown in Figure XIV.7 and the associated confidence intervals are given in Figure XIV.8. The point estimates the slope and intercept are seen to be very similar to those based on the threshold model fit in Figure XIV.3. In particular

	$\hat{\beta}_0$	$\hat{\beta}_1$	\hat{c}	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{c})$
Threshold Model	2.537	0.036	0.003	0.834	0.014	0.029
No Threshold Model	2.616	0.035	0 (by defn.)	0.257	0.006	0 (by defn.)

We see that the point estimates of β_0 , β_1 , c have not changed by much, but the standard errors have decreased markedly. Thus if there is no statistical evidence of background mortality we should eliminate it from the model to increase estimation precision.

Let's see how this affects the percentile point estimates and lower confidence bounds on them.

<u>Threshold Fit</u>			<u>No Threshold Fit</u>	
<u>Percentile</u>	<u>Point Estimate</u>	<u>Lower 97.5 Percent Confidence Bound</u>	<u>Point Estimate</u>	<u>Lower 97.5 Percent Confidence Bound</u>
1	3.773	-157.377	1.657	-17.646
3	16.086	-109.518	14.388	0.000
5	22.606	- 84.331	21.130	8.900
10	32.647	- 45.967	31.511	21.622
15	39.421	- 20.657	38.515	29.364
20	44.805	- 1.246	44.082	35.034
30	53.572	27.152	53.146	43.539
50	68.064	54.847	68.130	56.372
70	82.557	67.675	83.115	68.459
80	91.323	73.775	92.179	75.594

We see that the point estimates under the nonthreshold fit are slightly lower than the point estimates under the threshold fit in the lower portion of the curve.

However the increased precision of estimation under the nonthreshold fit results in substantial increases in the lower confidence bounds.

Suggestion:

1. If there is no observed response in the control group.

and 2. If the control group response rate, as estimated from the dose response fit - including threshold - is nonsignificant.

and 3. If there is no a priori reason to expect background rate

then, eliminate background threshold parameter from the model.

This raises in a conjecture: Suppose we fit a nonthreshold model even when a non zero background rate exists. We conjecture that the point estimates of nonthreshold response rates will be estimates of quantities lower than the true response rates. However the increased precisions of these estimates may well result in more accurate lower confidence bounds on the "true" response percentiles. This is a bias-variance trade off.

It is interesting to note that Charles Stephan [42], page 78 ff discusses Abbott's correction in connection with the estimation of LC-50 concentration in acute toxicity tests. He comments "...Abbott's formula ... is a statistically sound way of correcting for control mortality if, and only if, the cause of the control mortality does not make the rest of the test organisms more susceptible to the toxicant. This assumption is usually questionable in acute mortality tests with aquatic animals. ... If control mortality is below a specified minimum...it should be reported along with the results of the test, but correction of the LC50 for this mortality would seem to be a meaningless exercise. ...". It is interesting that we arrive at a similar suggestion, based on different reasoning. Our motivation is a bias-variance tradeoff.

The previous PROC PROBIT analyses on the DeFoe data treated concentration without any transformation. We also tried to fit a probit model using log concentration. Folklore states that a probit or logit fit will better fit the response vs logarithmic concentration relation than the response vs arithmetic concentration relation.

Finney [11] page 8-13 recommends using log concentration. Stephan [42] also recommends the use of a logarithmic transformation of concentration on a routine basis.

Finney, pages 9ff states "The frequency distribution of tolerances, as measured on the natural scale (i.e. arithmetic scale - P.F.) is usually markedly skew, but often a simple transformation of the scale of measurement will convert it to a distribution approximately of the familiar Gaussian or normal form ... normalization can often be effected by expressing the tolerances in terms of the logarithms of the concentrations instead of the absolute values. Indeed this transformation is now standard practice ... the justification is the widespread applicability of the normal distribution as an adequate approximation to the truth. ..."

Stephan, page 75, states "...Whenever any method is used to analyze concentration mortality data, whether or not a transformation such as probit, logit, or angle is used on the mortality data, the logarithmic transformation should probably be used on the concentration data. All of the methods assume that the concentration-mortality curve is linear, and it seems to be generally accepted that the curve is more likely to be linear if log concentration is used. ..."

We show by example of the DeFoe 1, 1, 2, trichloroethane data that the logarithmic transformation of concentration provides a much poorer fit of the probit model than does arithmetic concentration. The moral is that each time we fit a probit, logit, or other dose response model we should have an open mind as to using untransformed concentration, logarithmic concentration, or some other function of concentration. We should transform concentration in a manner suitable for each individual data set.

We first tried to fit the probit model with background response to logarithmic concentration. The attempted fit would not converge. To improve convergence performance we refitted the DeFoe data with logarithmic concentration using a specified background rate of 0 (the observed level). The output appears in Figures XIV.9. The maximum likelihood algorithm converges, however the resulting probit model is not an adequate fit to the data as indicated by the highly significant residual chi square statistic. (Chi square = 47.6774 with 4d.f.)

We break out the components of the chi square statistic by group, as discussed previously, to determine whether the very large chi square value is due just to one or two components with very small expected responses but with one or two observed responses. Such components can greatly distort the overall chi square.

$\log_{10}(\text{Conc})$	Trt Grp	O	N	\hat{p}	$N\hat{p}$	$N\hat{p}\hat{q}$	$\frac{(O - N\hat{p})^2}{N\hat{p}\hat{q}}$
-1.3098	1	0	40	0.0000	0.0000	0.0000	0.0000
0.29907	2	0	40	0.000015	0.00060	0.00060	0.00060
0.7764	3	2	40	0.0025428	0.10171	0.101455	35.518
1.1706	4	1	41	0.0476	1.9516	1.8587	0.4872
1.6840	5	9	40	0.4227	16.908	9.7610	6.4068
2.1673	6	40	40	0.884	35.36	4.102	5.249
							47.662

We see the third treatment group contributes the most to the statistic. It has a very small expected frequency and two observed responses. If this were the only large deviation between data and model, we might be inclined to consider the possibility of a reasonable probit fit with the exception of an outlier group. However, even if we disregard this group, the components of chi square from the remaining cells sum to 12.144 with 3d.f. This value is still significant at the 0.01 level, even after the largest component has been deleted. We thus conclude that the model does not fit the data well. This, coupled with probit plots suggests

the inappropriateness of the probit fit after a logarithmic transformation of concentration. The probit fit to untransformed concentration is superior in this case.

We now consider the Holcombe and Phipps compound D data. A listing of these data is contained in Figure XIV.10. The variable names correspond to those of the DeFoe data. The test consisted of six groups (control + 5 toxicant concentrations) and four tanks per group. Note that the control group concentration is 0 and there is a nonzero threshold response rate.

Probit models were fitted to the fry mortality data after pooling tanks within concentration groups. SAS PROC PROBIT was used to fit probit models both to concentration and to \log_{10} (concentration). These fits included background effects, to be fitted by maximum likelihood.

The probit fit vs untransformed concentration appears in Figures XIV.11, XIV.12. The residual chi square statistic is quite small (0.3361 with 3d.f.) signifying a good fit to the data. Figure XIV.12 contains the estimated percentiles of the probit response curve, adjusted for background, along with lower and upper 95 percent confidence limits calculated by use of Fieller's Theorem. For example for the 10th percentile the estimate for C_{10} is 78.77 while a lower 97.5 percent confidence bound is 58.72.

One difference between the fits to the DeFoe and to the Holcombe and Phipps data should be noted. In the DeFoe data no mortality was observed in the control group and the threshold response rate was estimated to be $\hat{c} = 0.003$ with an asymptotic standard error of 0.029. Thus there was no evidence of background mortality and we markedly improved precision of the fit by deleting the background correction.

In the case of the Holcombe and Phipps compound D data we observe $X = 6$ deaths within the control group, with each of the 4 tanks exhibiting at least one death. Thus we know that there is background variation. From our probit fit with arithmetic dosage we estimate $\hat{c} = 0.0718$ with an asymptotic standard error of 0.016. Thus \hat{c} is 4.5 asymptotic standard deviations from 0 and so is highly statistically significant.

We now fit a probit model to the same data using \log_{10} (concentration). The estimated parameter values, their estimated asymptotic variance-covariance matrix, and the residual chi square statistic appear in Figure XIV.13. The residual chi square statistic is 0.5046 with 3d.f., which is very small, thus indicating a good fit to the data¹. We thus have

¹ Note that the chi square value 0.2287, given by SAS in Figure XIV.13 is incorrect in this case. It seems to be omitting the control group contribution to the chi square statistic. This problem has been brought to the attention of the program developer and has been corrected in later versions of the program.

good probit fits to the data using both arithmetic and logarithmic concentration.

The estimated background level is $\hat{c} = 0.0738$ with a standard error of 0.0156. Thus \hat{c} is 4.73 standard deviations from 0 and so is highly statistically significant.

The estimated percentiles (after adjusting for control group mortality) and 95 percent confidence intervals (by Fieller's theorem) are shown in Figure XIV.14. The lower confidence bound is a 97.5 percent one sided bound. This display is analogous to that in Figure XIV.12.

We have thus fitted two distinct models which seem to fit the data well: the probit model with arithmetic concentration (Figure XIV.11) and the probit model with logarithmic concentration (Figure XIV.13). The parameter estimates associated with these two fits are somewhat different. Namely

	<u>Arithmetic (StdErr)</u>	<u>Logarithmic (StdErr)</u>
$\hat{\mu}$	109.916 (3.638)	2.028 (0.016)
$\hat{\sigma}$	24.300 (3.861)	0.1044(0.015)
\hat{c}	0.0718(0.016)	0.0738(0.016)

We see that the estimated background levels are somewhat similar, however the estimates $\hat{\mu}$, $\hat{\sigma}$ are very different.

Even though the parameter estimates differ considerably, the model fits may still be very similar. We compare the estimated response distribution percentiles and associated lower confidence bounds in Figures XIV.12, XIV.14 for the arithmetic and logarithmic concentration fits respectively. These are shown below.

<u>Arithmetic (Figure XIV.12)</u>			<u>Logarithmic (Figure XIV.14)</u>	
<u>Percentile</u>	<u>Estimate</u>	<u>Lower 97.5% Conf. Bound</u>	<u>Estimate</u>	<u>Lower 97.5 % Conf Bound</u>
1	53.39	22.40	60.97	45.76
2	60.01	31.92	65.10	50.13
3	64.21	37.95	67.87	53.12
5	69.95	46.15	71.83	57.47
10	78.77	58.72	78.38	64.83
15	84.73	67.13	83.14	70.29
20	89.46	73.77	87.12	74.92
30	97.17	84.43	94.03	83.01

The point estimates of the response distribution percentiles corresponding to the arithmetic and logarithmic fits are quite similar beyond

the third percentile. However there is considerable discrepancy between corresponding lower confidence bounds on safe dose, based on each of the two fits -- below the 10th percentile! Below the third percentile the discrepancy is fifty percent or more.

How can we choose between the two fits?

1. Prior knowledge or mechanistic information

None here. Since the probit model is an empirical model, no much in the way of mechanistic arguments will distinguish between the two fits.

2. Magnitude of residual chi square

Arithmetic fit Residual chi square = 0.336 with 3d.f.
($\alpha = 0.953$)

Logarithmic fit Residual chi square = 0.505 with 3 d.f.
($\alpha = 0.918$)

Both chi square values are quite small and the question of which one is larger is probably just a matter of chance fluctuations. Therefore we should not use these two statistically insignificant values to distinguish between the fits.

3. Appearances of plots of predicted and observed responses

Scatterplots of predicted and observed responses vs arithmetic concentration are shown in Figures XIV.15, XIV.16. Similar plots vs logarithmic concentration are shown in Figures XIV.17, XIV.18. The probit plots (Figures XIV.15, XIV.17) indicate greater discrepancies between observed and predicted responses (after adjusting both for background) at the low percentiles of the logarithmic concentration fit than of the arithmetic concentration fit. Similarly at the highest treatment group. Thus the arithmetic concentration fit seems to be a (slightly) better approximation to the data at the low percentiles than does the logarithmic concentration fit.

4. Conservativeness. Below the 25th percentile the lower confidence bounds based on the arithmetic concentration fit are lower than those based on the logarithmic concentration fit. The discrepancy is especially noticeable for the low percentiles, in particular below the 10th percentile. Above the 10th percentile both lower bounds are similar. Thus the arithmetic concentration fit seems to be more conservative than the logarithmic concentration fit at the low percentile.

Opinion I would prefer the arithmetic concentration fit in this case. However further experimentation at the low concentrations would be needed to distinguish between the differing conclusions at the low percentiles.

Alternative analysis ignoring background

We remarked with respect to the analysis of the DeFoe data that an alternative way of handling background mortality is to ignore it. The hope is that the improved precision will offset the downward bias and result in higher values for lower confidence bounds on safe dose.

Note: If the control group response rate is significantly different from 0, as is the case with the Holcombe and Phipps data, we would not expect the dose response fit ignoring background to be a good fit to the data. Probit models ignoring background response, fitted to both arithmetic and logarithmic concentrations, show large and highly significant residual chi square statistics 28.96 and 65.03 respectively with 4d.f. The plots of fitted and observed responses vs concentration also show discrepancies.

We now compare the percentile point estimates and their lower confidence bounds based on the fit ignoring background response with those based on the fit including background response. Comparisons pertain to the fit vs untransformed concentration.

<u>Background Included</u>			<u>Background Excluded</u>	
<u>Percentile</u>	<u>Point Estimate</u>	<u>Lower 97.5% Conf. Bound</u>	<u>Point Estimate</u>	<u>Lower 97.5% Conf. Bound</u>
1	53.39	22.403	-7.48	-24.234
3	64.21	37.946	13.37	0.648
5	69.95	46.147	24.41	13.197
7	74.05	52.006	32.33	21.991
10	78.77	58.715	41.42	32.626
15	84.73	67.135	52.89	45.860
20	89.46	73.775	62.01	56.369
25	93.53	79.421	69.83	64.178
50	109.92	101.288	101.40	97.761

We see that in this example, with the background level many standard deviations from 0, the bias-variance trade off is such that it does not pay to reduce the assumed background response level to 0 in order to lessen the standard deviation.

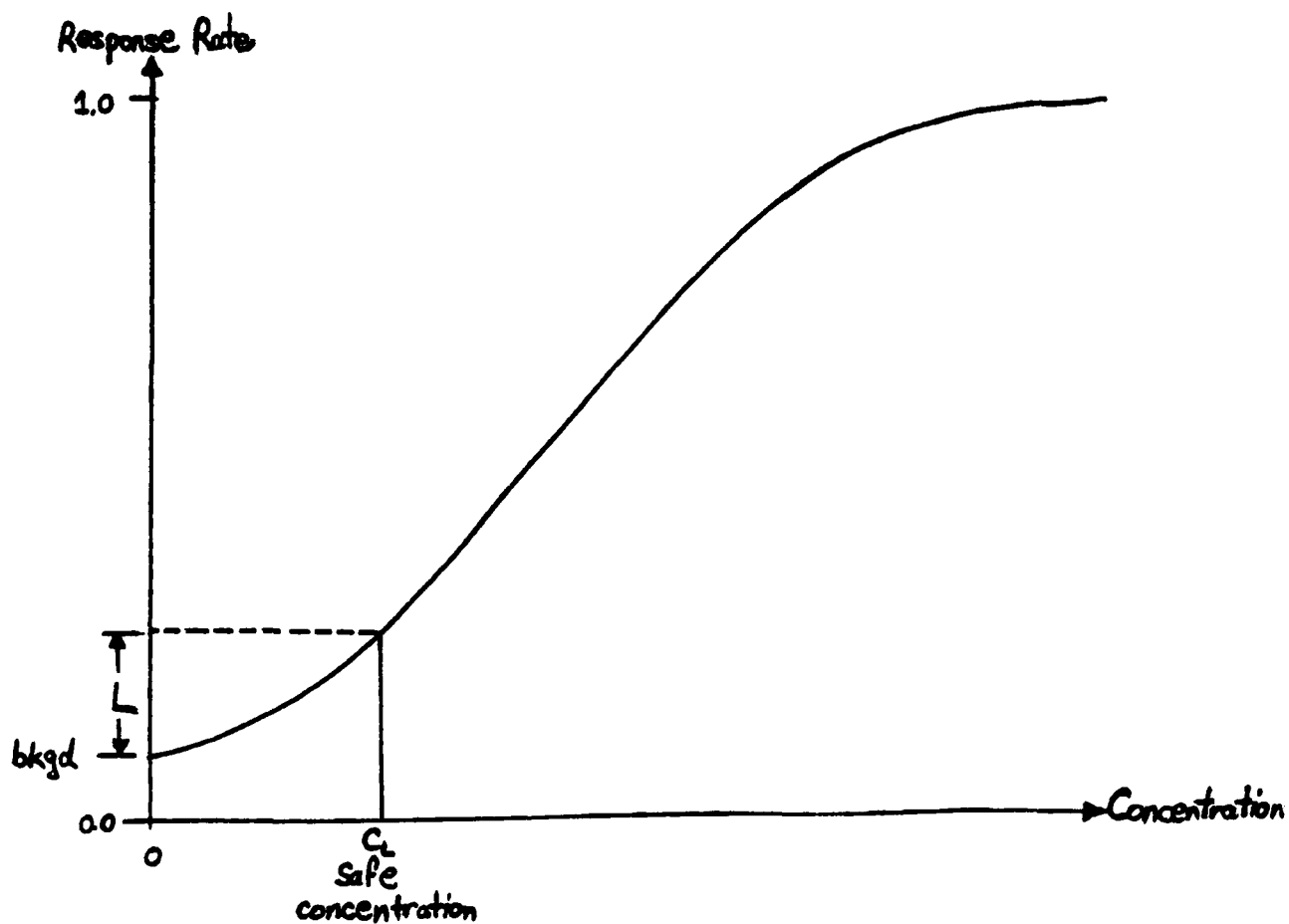


Figure XIV.1 Schematic representation of dose response curve formulation

Basic Data for Acute Analysis - Dose Compound C

COMBINED DEFOLI DATA - SURVIVAL AND TOXICANT										14:26 SUNDAY, AUGUST 3, 1980		1
URS	TRT	CONCMEAN	REP	EMHSUM	FRYSUM	DEADESUM	DEAFESUM	PRPDEADE	PRPDEADF	LOGCONC	(log ₁₀ Conc)	
1	1	0.049	A	50	20	22	0	0.44	0.00	-1.3059		
2	1	0.049	A	50	20	19	0	0.38	0.00	-1.3059		
3	2	1.491	A	50	20	22	0	0.38	0.00	0.2590		
4	2	1.491	A	50	20	19	0	0.38	0.00	0.2590		
5	3	5.976	A	50	20	20	0	0.40	0.00	0.7764		
6	3	5.976	A	50	20	17	2	0.34	0.10	0.7764		
7	4	14.812	A	50	21	16	1	0.32	0.00	1.1706		
8	4	14.812	A	50	20	21	1	0.44	0.05	1.1706		
9	5	48.307	A	50	20	22	4	0.34	0.25	1.6840		
10	5	48.307	A	50	20	17	5	0.38	0.25	1.6840		
11	6	146.984	A	50	20	19	20	0.38	1.00	2.1673		
12	6	146.984	A	50	20	19	20	0.38	1.00	2.1673		

Figure XIV.2 Basic data

DeFoe Compound C

PROBIT FIT OF FRY DATA: OPTIC, C=0

11:34 THURSDAY, JULY 29, 1980

2

PROBIT ANALYSIS ON CONCENTR

ITERATION	INTERCEPT	SLOPE	C	MU	SIGMA
0	2.06207915	0.075222552	0	80.78807155	39.64239896
1	2.98228666	0.028993502	0	69.30563326	36.58606093
2	2.91835560	0.03106586	0	66.58751373	37.18939980
3	2.95487504	0.03132432	0	67.13497720	31.32409521
4	2.70157399	0.03264760	0	67.63356893	30.6301158
5	2.72778375	0.03344106	0	67.94601907	29.30264106
6	2.66095331	0.03433231	0	68.12747337	29.12622930
7	2.58033205	0.03537571	0	68.17157183	29.26799156
8	2.54168422	0.03610287	0.00143209	68.09279850	27.69882527
9	2.53730962	0.03618162	0.00301437	68.06467395	27.63833995
10	2.53714566	0.03618433	0.00308614	68.06412265	27.63627618
11	2.53714560	0.03619433	0.00308614	68.06412225	27.63627534

COVARIANCE MATRIX					COVARIANCE MATRIX				
INTERCEPT	SLOPE	C	MU	SIGMA	INTERCEPT	SLOPE	C	MU	SIGMA
0.60509113	-0.01099208	-0.02241393	58.76577025	40.34127025	0.60509113	-0.01099208	-0.02241393	58.76577025	40.34127025
-0.01099308	0.00018664	0.00046624	40.34127025	110.59470236	-0.01099308	0.00018664	0.00046624	40.34127025	110.59470236
-0.02241353	0.00034624	0.00083146	-0.03185415	-0.26444583	-0.02241353	0.00034624	0.00083146	-0.03185415	-0.26444583

CHI-SQ = 4.2108 WITH 3 DF PROB > CHI-SQ = 0.2396

NOTE: SINCE THE CHI-SQUARE IS SMALL (P > 0.10), FINANCIAL LIMITS WILL BE COMPUTED USING A T VALUE OF 1.96.

Figure XIV.3 Output from PROC PROBIT fit to DeFoe fry mortality data -- arithmetic concentration

DeFoe Compound C

PROBIT FIT OF FRY DATA: OPTIC, C=0
 PROBIT ANALYSIS OF CONCENTRATIONS

11:34 TUESDAY, JULY 29, 1980 3

⑤

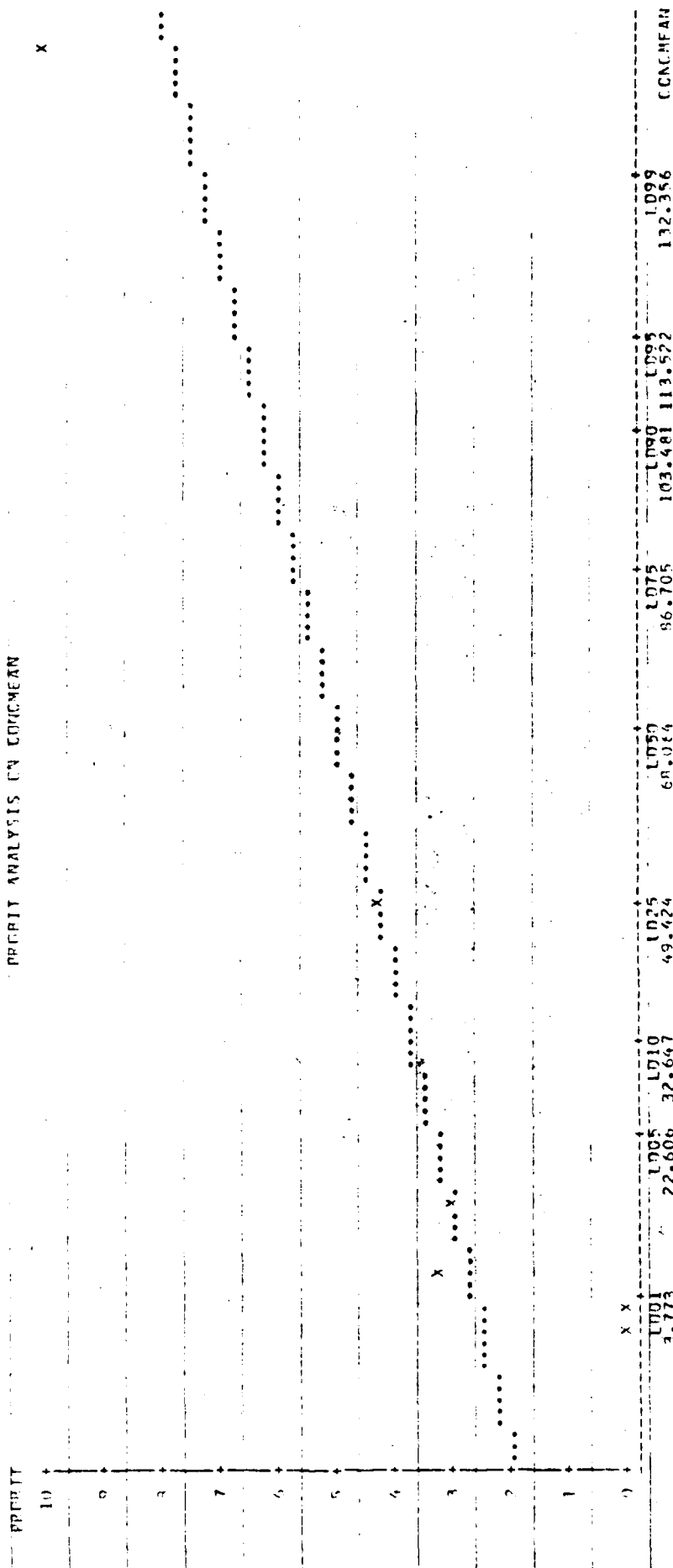


Figure XIV.4 Output from PROC PROBIT fit to DeFoe fry mortality data -- arithmetic concentration

DeFoe Compound C

PROBIT FIT OF FRY DATA: DPTC, C=0
 PROBIT ANALYSIS ON CUMMEAN

11:24 TUESDAY, JULY 29, 1980 4

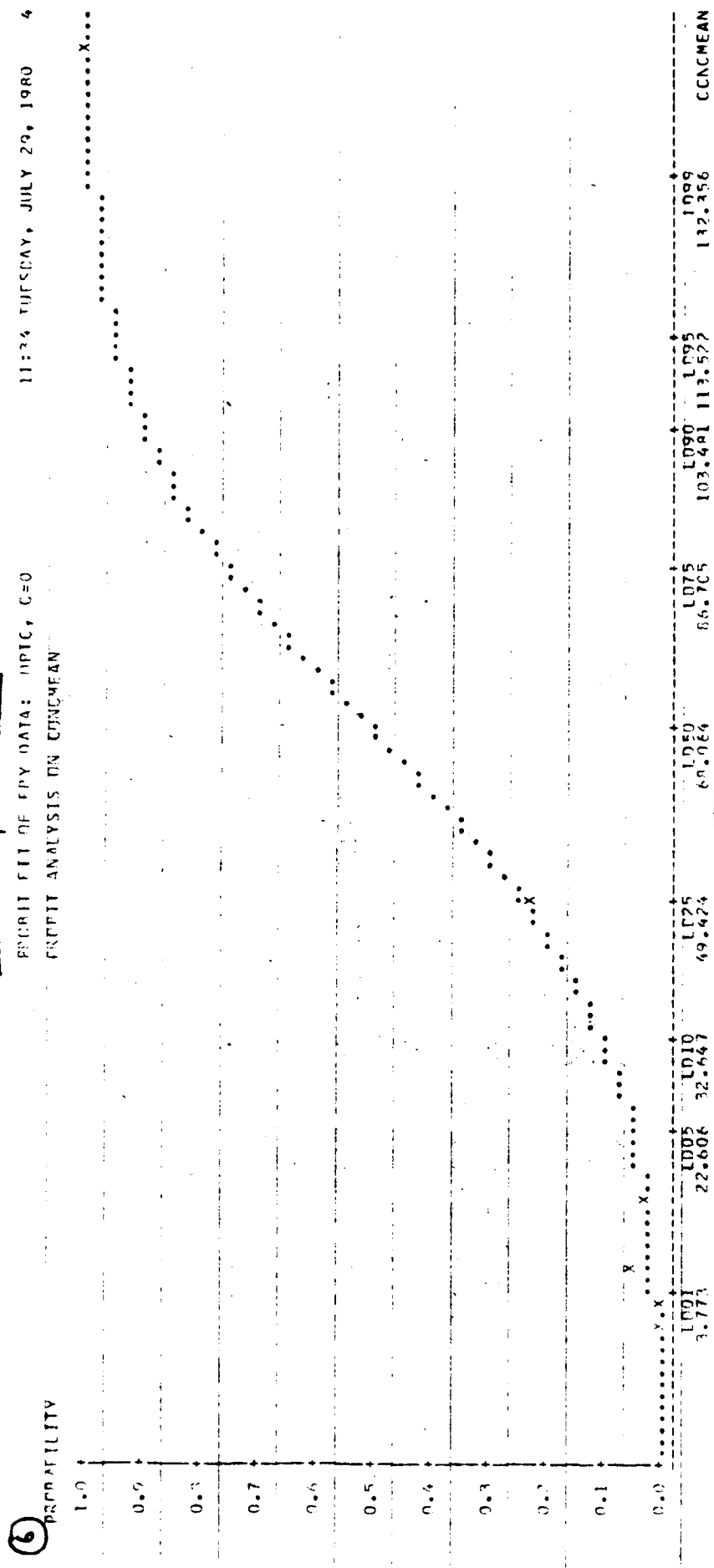


Figure XIV.5 Output from PROC PROBIT fit to DeFoe fry mortality data -- arithmetic concentration

DeFoe Compound C

PROCIT FIT OF FRY DATA: PROC, C=0

11:34 TUESDAY, JULY 29, 1980

5

(7)

PROBIT ANALYSIS ON CONCENTR

PROBABILITY	CONCEN	95 PERCENT FIDUCIAL LIMITS
		LOWER UPPER
0.01	3.77253109	-157.37703054 29.00105204
0.02	11.30615198	-129.06089974 33.64097224
0.03	16.08599233	-109.51837857 36.64247678
0.04	19.69168004	-95.61015631 38.94122745
0.05	22.60649411	-84.33072910 40.84448667
0.06	25.09597119	-74.75938750 42.49388327
0.07	27.27875513	-66.70431880 43.86722878
0.08	29.253317781	-58.73027935 45.31227803
0.09	31.01064705	-52.16719618 46.56072646
0.10	32.64681039	-45.96713818 47.73486681
0.15	39.82058218	-20.65757158 52.95866691
0.20	44.80484615	-1.24640993 57.31410483
0.25	49.42373784	14.43687889 62.94940285
0.30	53.57164532	27.15181500 68.93020115
0.35	57.41529577	37.15115102 76.25524847
0.40	61.06255197	44.71983280 85.12571419
0.45	64.59131068	50.40837319 95.34163790
0.50	69.06412229	54.84680955 106.56666684
0.55	71.53693383	59.52519224 118.52931635
0.60	75.06569274	61.75719105 131.0528563
0.65	78.71294474	64.75757191 144.84304117
0.70	82.55659918	67.67468638 159.05031086
0.75	86.70450667	70.61423276 174.78645888
0.80	91.32339595	73.77499874 192.56931976
0.85	96.70728072	77.29757321 213.20835084
0.90	103.48143412	81.59019088 239.44866407
0.91	105.11759766	82.61004052 245.80344258
0.92	106.89505679	83.71176133 252.11305308
0.93	108.84948538	84.91708488 260.31703800
0.94	111.03227333	86.25582405 268.81665164
0.95	113.52174900	87.77445942 278.51879860
0.96	116.44655666	89.56887079 289.02770745
0.97	120.04225217	91.71783918 303.98507584
0.98	124.87209232	94.58330619 322.64363409
0.99	132.55571252	99.06649286 352.11595912

Figure XIV.6 Output from PROC PROBIT fit to DeFoe fry mortality data -- arithmetic concentration

DeFoe Compound C

Assumption of 1.41m/1.1h

PROBIT FIT OF FRY DATA C=0

14:12 SUNDAY, AUGUST 3, 1980 2

PROBIT ANALYSIS ON CONCMEAN

ITERATION	INTERCEPT	SLOPE	MU	SIGMA
0	2.96207915	0.02522552	80.78807155	39.64239896
1	2.74160642	0.03025002	73.53118048	32.61683378
2	2.55083553	0.03378769	68.51258369	29.58782490
3	2.61372376	0.03489489	68.24160141	28.65756615
4	2.61569724	0.03499552	68.13102397	28.57476080
5	2.61566871	0.03499685	68.13023421	28.57415963
6	2.61566871	0.03499685	68.13023417	28.57415960

C = 0.00000000

COVARIANCE MATRIX		COVARIANCE MATRIX	
INTERCEPT	SLOPE	MU	SIGMA
0.06400365	-0.00118564	55.25332279	28.22800846
-0.00118544	0.00003515	28.22800846	23.43827242

CHI-SQ = 4.4131 WITH 4 DF PROR > CHI-SQ = 0.3530

NOTE: SINCE THE CHI-SQUARE IS SMALL (P > 0.10), FIDUCIAL LIMITS WILL BE COMPUTED USING A T VALUE OF 1.96 .

Figure XIV.7 PROC PROBIT fit to DeFoe fry mortality data --- no background correction

DeFoe, Compound C
Assumption of No Threshold

PROBIT FIT OF FRY DATA C=0				14:12 SUNDAY, AUGUST 3, 1980		5	
PROBIT ANALYSIS ON CONCMEAN							
PROBABILITY	CONCMEAN	95 PERCENT FIDUCIAL LIMITS					
		LOWER	UPPER				
0.01	1.65679804	-17.6459257	13.01179939				
0.02	0.44408512	-0.74721373	19.62281154				
0.03	1.05413753	-0.30039376	13.98522082				
0.04	1.10541104	4.95710741	27.39392772				
0.05	1.12603219	9.89665113	32.74622421				
0.06	1.17043501	13.17943501	35.00749196				
0.07	2.26074559	15.99131324	37.08713550				
0.08	2.29819252	17.45410685	39.02628529				
0.09	2.15108752	18.64611647	40.85314988				
0.10	3.85150213	21.62190899	48.85597169				
0.15	4.48081617	25.36324454	55.69817433				
0.20	4.85725643	30.03441022	61.84664782				
0.30	5.71453302	43.53811468	67.56952322				
0.40	6.61201562	47.05659004	75.98545373				
0.50	6.53936559	50.30739009	78.21106275				
0.60	6.13023417	53.38691837	83.31100829				
0.70	7.11205014	55.31487666	88.41955265				
0.80	7.95494172	62.2709587	93.54832116				
0.90	8.21145300	68.29697435	98.79347127				
0.95	8.74032191	68.45944190	104.24417937				
0.97	9.17365721	71.84701069	110.01505369				
0.98	9.77456308	75.52405523	116.26797137				
0.99	10.44182242	78.34333627	123.25607137				
0.995	10.82789312	80.67977569	131.74400845				
0.998	11.05665253	85.08313222	144.24002825				
0.999	11.225665415	88.64519222	146.95335922				
0.9995	11.513054415	93.36002769	153.27517138				
0.9998	11.815461730	95.68067332	157.09307358				
0.9999	12.133333333	98.52921009	161.56020103				
0.99995	12.61433324	102.30908631	174.39805137				
0.99998	13.6036690	108.25342671	185.96320395				

Figure XIV.8 PROC PROBIT fit to DeFoe fry mortality data -- no background correction

DeFoe Compound C---No threshold

PROBIT FIT OF FRY DATA: C=0, LOG10
 PROBIT ANALYSIS ON LOG10(CONCMEAN)

14:12 SUNDAY, AUGUST 3, 1980 6

ITERATION	INTERCEPT	SLOPE	MU	SIGMA
0	2.2742663	1.04829111	2.40016645	0.95393349
1	0.96478358	2.22661144	1.90416515	0.44710293
2	0.14728138	2.76487799	1.75512939	0.34169661
3	-0.04577939	2.89771407	1.75192761	0.34858824
4	-0.03130657	2.87198079	1.75186075	0.34819174
5	-0.03140211	2.87198394	1.75186075	0.34819136

C = 0.00000000

COVARIANCE MATRIX		COVARIANCE MATRIX	
INTERCEPT	SLOPE	MU	SIGMA
4.79006845	-2.71778779	0.03144517	0.00555591
-2.71778779	1.62664212	0.00555591	0.02390664

CHI-SQ = 47.5775 WITH 4 DF

PROB > CHI-SQ = 0.0001

NOTE: THE ABOVE VARIANCES AND COVARIANCES HAVE BEEN MULTIPLIED BY THE HETEROGENEITY FACTOR H= 11.9193681.
 CHECK TO BE SURE THAT THE LARG CHI-SQUARE (P < 0.10) IS NOT CAUSED BY SYSTEMATIC DEVIATION FROM THE MODEL.
 IF VALUES OF 2.78 WILL BE USED IN COMPUTING EDUCIAL LIMITS.

(t(975H)).

strong statistical evidence of lack of fit

Figure XIV.9 PROC PROBIT fit to DeFoe fry mortality data -- logarithmic transformation of concentration

COMBINED HULF:16 DATA - SURVIVAL AND TOXICANT

14:31 SUNDAY, AUGUST 3, 1980 1

[illegible]

Figure XIV.10 Basic data

ORIGIN: FLY 16 40Y DATA: 0010, 000, 00

THE JOURNAL OF THE

ITERATION	INTERCEPT	SLOPE	C	"U	SIGMA
7	1.871176691	0.07994454	0.060909390	102.081752235	374.42513657
8	1.041383478	0.07327763	0.07930192	108.48087186	375.52701138
9	0.63208492	0.06931894	0.07668974	109.34007066	275.04241158
10	0.49274193	0.06122522	0.07165283	109.84691308	277.37231790
11	0.476737495	0.05115032	0.07176560	107.81611911	274.47120525
12	0.47675929	0.04115169	0.07176700	109.34269245	274.30134059
13	0.47675923	0.03115169	0.07176700	109.91629255	274.30034086

```
CHI-SQ = 0.3361 WITH 3 DF      PROB > CHI-SQ = 0.9531
```

NOTE: SINCE THE CHI-SQUARE IS SMALL ($P > 0.10$), FINANCIAL LIMITS WILL BE COMPUTED USING A T VALUE OF 1.06.

Figure XIV.11 Output from PROC PROBIT fit to Holcombe and Phipps fry mortality data -- arithmetic concentration

AD-A134 309

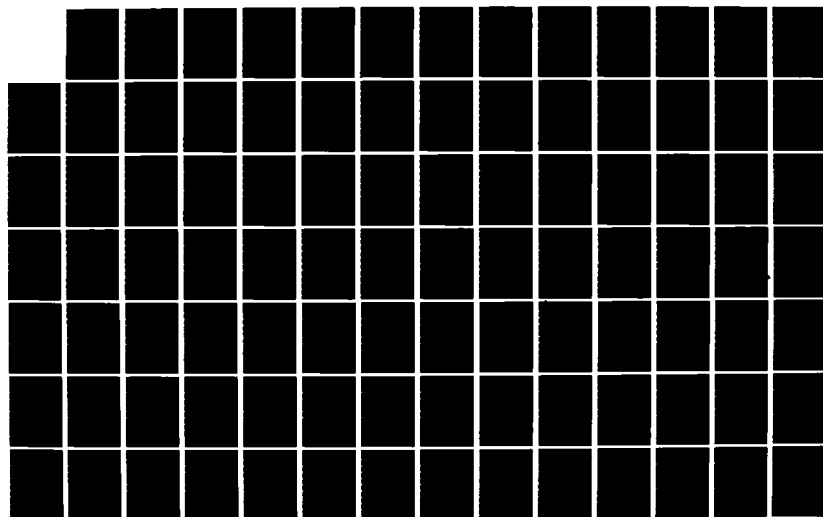
DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF
TOXICITY(U) OHIO STATE UNIV RESEARCH FOUNDATION
COLUMBUS P I FEDER ET AL MAY 81 DAMD17-79-C-9150

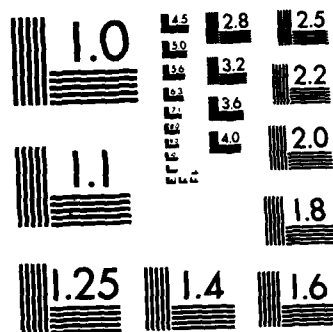
3/5

UNCLASSIFIED

F/G 6/20

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Holcombe and Phipps-Compound D-Arithmetic Concentration

PROBIT FIT OF FRY DATA: UPTC, C=0.06

10:48 TUESDAY, JULY 29, 1980 5

PROBABILITY	CONC-MEAN	95 PERCENT FIDUCIAL LIMITS	
		LOWER	UPPER
0.01	53.38524434	22.40313687	70.11953833
0.02	60.00949506	31.91063088	75.57264847
0.03	64.21236688	37.97561854	78.55725260
0.04	67.37302435	42.97133093	81.03415418
0.05	69.94578881	46.14701499	83.05444997
0.06	72.13476430	49.17108417	84.77864582
0.07	74.05406758	52.00644015	86.29422647
0.08	75.77257475	54.55223977	87.65463358
0.09	77.33141523	56.67353509	88.89493101
0.10	78.77141523	58.71542039	90.03544745
0.15	86.77306095	67.13481108	94.81261155
0.20	90.62550173	73.77676220	98.55768155
0.25	93.52576114	79.42082118	102.00694101
0.30	96.17318116	84.43787248	105.06376805
0.35	103.55287189	89.02774578	107.56335671
0.40	103.75987160	93.31492739	110.71914009
0.45	106.86267199	97.38227347	113.53399483
0.50	109.51629255	101.28828228	116.44122569
0.55	111.60727115	105.07697391	119.41578273
0.60	113.21971121	108.78513118	122.57982307
0.65	114.26594037	112.44979137	126.01813207
0.70	115.30662334	116.11802471	129.83535194
0.75	116.36797377	119.86119668	134.17016507
0.80	117.40197715	123.79717289	139.22882512
0.85	118.40584323	128.14177819	145.39407258
0.90	119.38433237	133.34008548	153.82772668
0.91	142.49709683	134.56285717	155.72776669
0.92	144.06001633	135.87978590	157.47776669
0.93	145.77851752	137.31558877	159.84277055
0.94	147.68782080	138.90582964	162.50346651
0.95	149.88679639	140.70459194	165.55299550
0.96	152.45853003	142.80037534	169.15529950
0.97	155.42021964	145.35577601	173.60030879
0.98	158.82233964	148.72229388	179.54181301
0.99	168.44713916	153.9783385	188.65802189

Figure XIV.12 Output from PROC PROBIT fit to Holcombe and Phipps fry mortality data -- arithmetic concentration

Holcombe and Phipps Compound D-Logarithmic Concentration

PROBIT FIT OF FRY DATA: OPTC, C=0.06, LOG10

10:48 TUESDAY, JULY 29, 1980

6

PROBIT ANALYSIS ON LOG10(CONCENTRATION)

ITERATION	INTERCEPT	SLOPE	C	MU	SIGMA
0	-3.38255783	4.07936591	3.06000000	2.05486785	0.24513614
1	-7.53934763	6.20214281	0.06711886	2.02172830	0.16123460
2	-11.30357495	8.06132485	0.07078098	2.02231823	0.14904140
3	-15.54557145	9.15780456	0.07226138	2.02310667	0.10915647
4	-19.30181480	9.2342442	0.07364437	2.02350746	0.10500425
5	-22.42476950	9.27838983	0.07383784	2.02397847	0.10440168
6	-25.42715117	9.27952086	0.07383425	2.02398777	0.10438935
7	-28.42716561	9.27952761	0.07383444	2.02398785	0.10438928
8	-31.42716590	9.27952727	0.07383443	2.02398785	0.10438928
9	-34.42716594	9.27952729	0.07383443	2.02398785	0.10438928
10	-37.42716593	9.27952729	0.07383443	2.02398785	0.10438928

COVARIANCE MATRIX		COVARIANCE MATRIX	
INTERCEPT	SLOPE	MU	SIGMA
8.71918018	-4.19505292	0.00025831	-0.00011681
-4.19505292	2.01301515	-0.00011681	0.00023904
-0.01663896	0.00776935	0.00009216	-0.00008466
			0.00008466

CHI-SQ = 0.97 WITH 3 DF PROB > CHI-SQ = 0.9728

NOTE: SINCE THE CHI-SQUARE IS SMALL (P > 0.10), FURTHER LIMITS WILL BE COMPUTED USING A T-VALUE OF 1.96.

~~0.97 WITH~~
~~3 DF~~
~~PROB > CHI-SQ = 0.9728~~
~~(See text)~~

Figure XIV.13 Output from PROC PROBIT fit to Holcombe and Phipps fry mortality data — logarithmic concentration

Holcombe and Phipps-Compound D-Logarithmic Concentration

PROBIT FIT OF FRY DATA: OPTIC, C=0.06, LOG10

10:48 TUESDAY, JULY 29, 1980

9

PROBIT ANALYSIS ON LOG10(CONCMEAN)

PROBABILITY	LOG10(CONCMEAN)	95 PERCENT FIDUCIAL LIMITS LOWER	CONCMEAN	95 PERCENT FIDUCIAL LIMITS UPPER
0.01	1.78514208	1.66052017	60.97363102	45.76359871
0.02	1.81359347	1.69511242	65.10270036	50.13792007
0.03	1.83165315	1.72412651	67.86643009	52.12165390
0.04	1.84528276	1.7445741	70.02207391	53.47873014
0.05	1.85668615	1.75944079	71.82617760	55.46994043
0.06	1.86568615	1.77245741	73.39882433	57.21850065
0.07	1.87393108	1.78385500	74.8057777	59.79315989
0.08	1.88131343	1.79404667	76.08732079	62.23671131
0.09	1.88802739	1.80330344	77.27293172	63.57749120
0.10	1.89420750	1.81181314	78.38021198	65.83554128
0.15	1.91979531	1.84691062	83.13218364	70.29274374
0.20	1.94017161	1.88091362	87.1225691	72.92162289
0.25	1.95757835	1.90817905	90.69358709	79.10046353
0.30	1.97324605	1.93837317	94.022199316	86.77071470
0.35	1.98776452	1.96372422	100.35598757	90.44247120
0.40	2.00154113	1.99350822	103.48337151	94.08235358
0.45	2.01487015	2.02442956	106.65462774	97.73321806
0.50	2.02798785	2.05690583	109.92350514	101.21899567
0.55	2.04110457	2.08905455	113.30681987	105.13471402
0.60	2.05443477	2.12126815	117.08467446	109.12384134
0.65	2.06821118	2.15353523	120.98426400	113.2049849
0.70	2.08229644	2.1858155	125.42892400	117.6304849
0.75	2.09837735	2.2180810	130.5700511	122.4359918
0.80	2.11384409	2.2503454	136.82930413	127.9959148
0.85	2.131618039	2.2826109	143.21336435	133.9647120
0.90	2.151794810	2.3148764	147.21332656	139.66122124
0.91	2.161794810	2.3248764	149.50735326	140.50350012
0.92	2.17466226	2.3348764	152.07037517	142.83235012
0.93	2.182044655	2.3448764	154.98458336	145.46445582
0.94	2.19028955	2.3548764	158.37730226	148.46445582
0.95	2.19969294	2.3648764	162.41837041	151.88991501
0.96	2.20707495	2.3748764	167.41837041	155.88991501
0.97	2.21432253	2.3848764	174.73332285	159.88991501
0.98	2.22432253	2.3948764	186.56686866	166.37953309
0.99	2.27083164	2.4448764		186.37953309

Figure XIV.14 Output from PROC PROBIT fit to Holcombe and Phipps fry mortality data -- logarithmic concentration

Holcombe and Phipps-Compound D-Arithmetic Concentration

PROBIT FIT OF FRY DATA: DPTC, C=0.06

10:49 THURSDAY, JULY 29, 1980 3

PROBIT ANALYSIS ON CONCMEAN

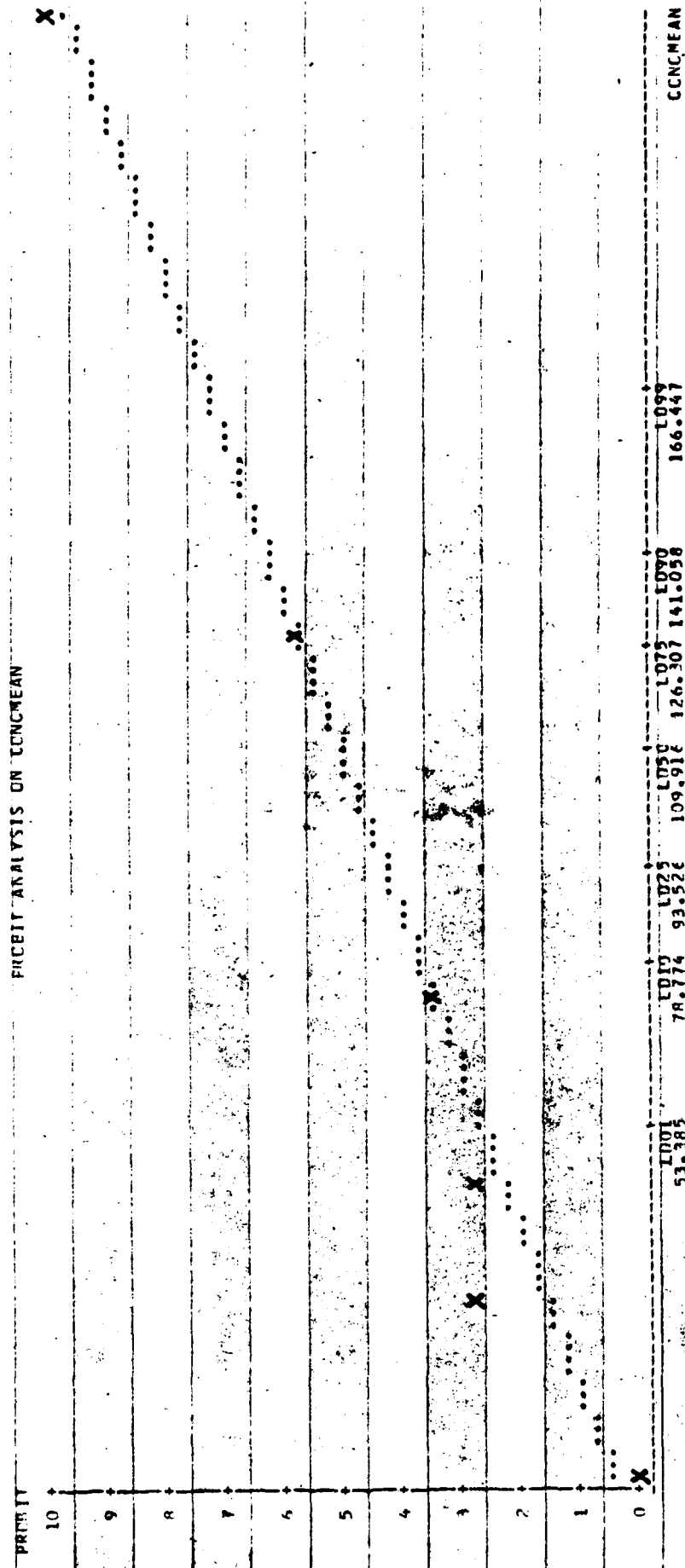


Figure XIV.15 Plots of observed and predicted based on probit fit -- arithmetic concentration

Helicobacter Phages - Concentrated D-Arithmetic Concentration

PROBIT FIT OF FRY DATA: OPTIC, C=0.06

10:48 TUESDAY, JULY 29, 1980 4

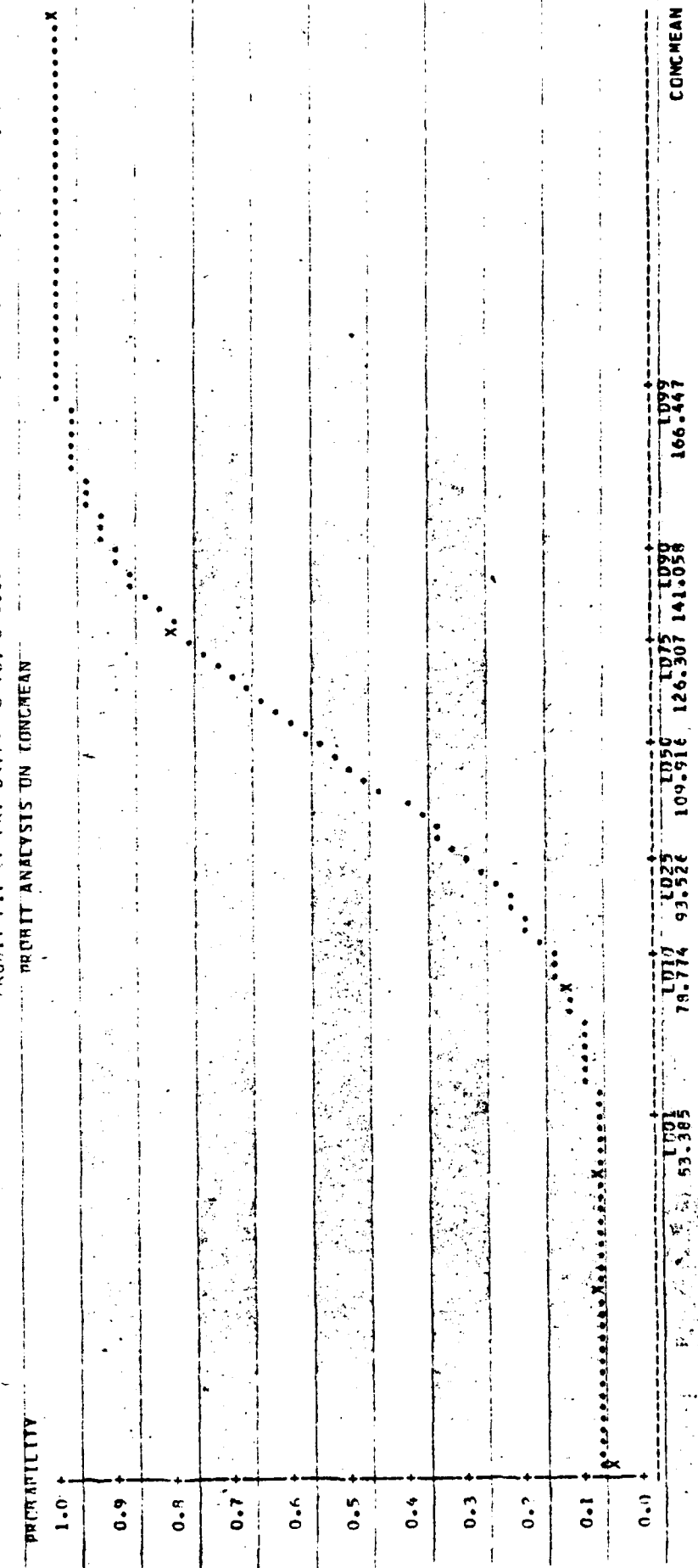


Figure XIV.16 Plots of observed and predicted based on probit fit --- arithmetic concentration

Atkins and Phillips - Compound D - Logarithmic Concentration

PROBIT FIT OF FRY DATA: OPTC, C=0.06, LOG10

10:49 TUESDAY, JULY 29, 1980 7

PROBIT ANALYSIS ON LOG10(CONC MEAN)

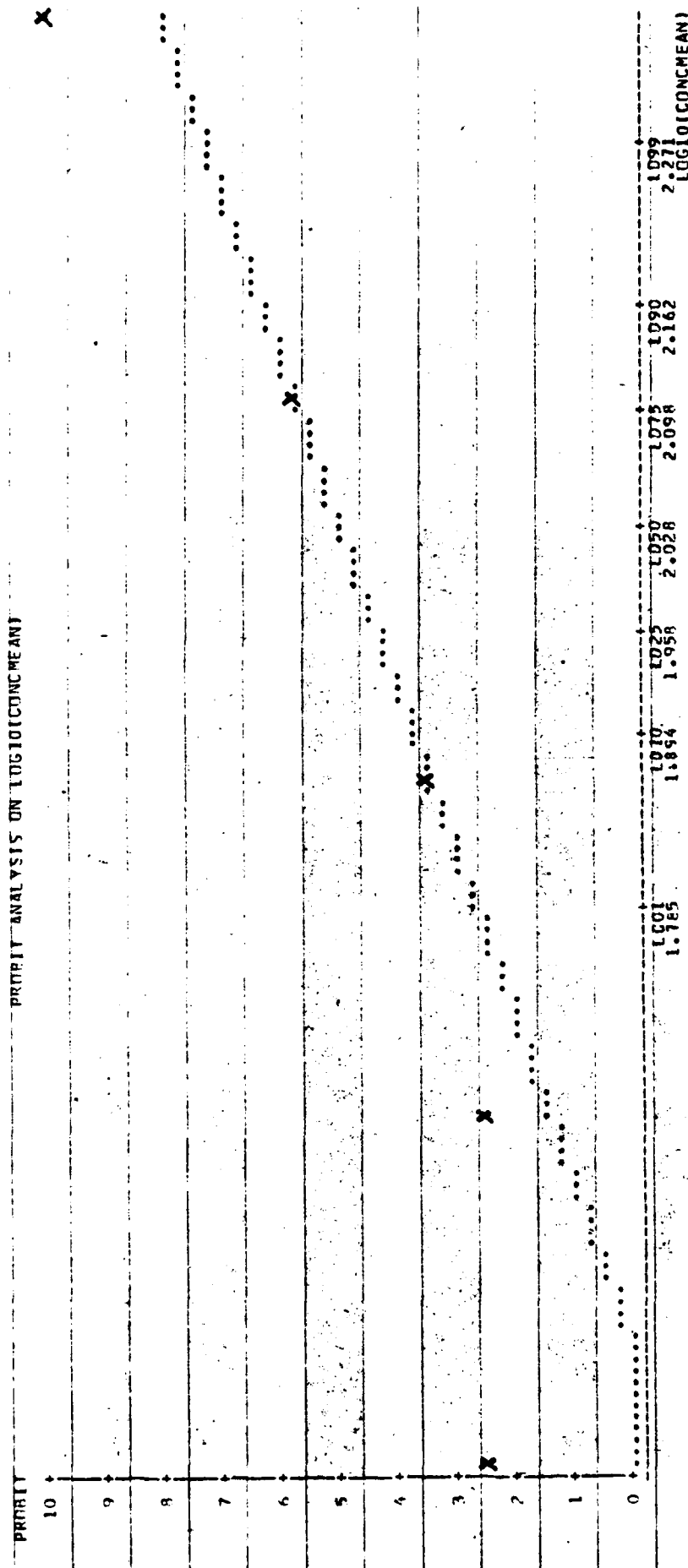


Figure XIV.17 Plots of observed and predicted based on probit fit -- logarithmic concentration

10:48 TUESDAY, JULY 25, 1980 8

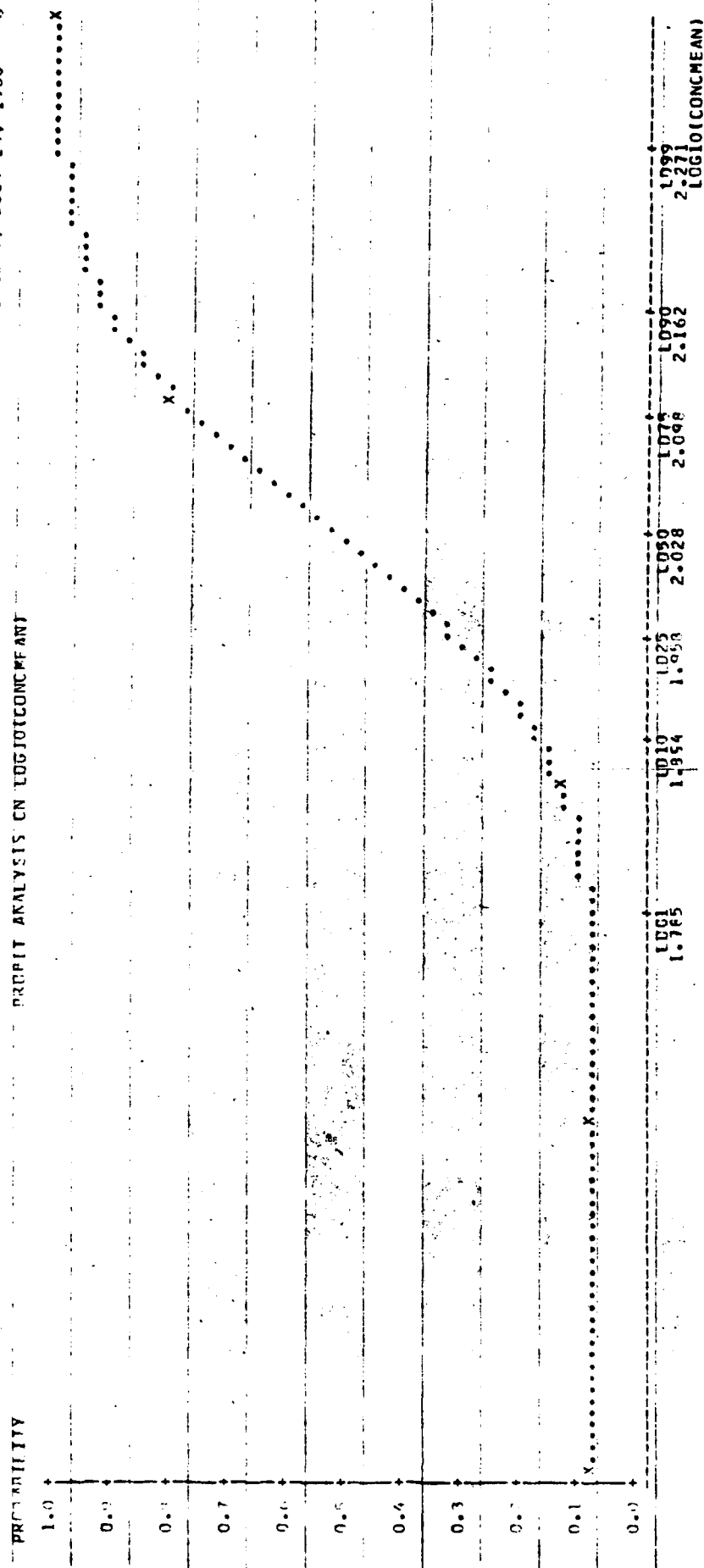


Figure XIV.18 Plots of observed and predicted based on probit fit --- logarithmic concentration

XV. DOSE RESPONSE CURVE ESTIMATION -- MAXIMUM LIKELIHOOD ESTIMATION BY NONLINEAR LEAST SQUARES REGRESSION

We have seen in the previous section how standard dose response models can be fitted to the data by use of SAS PROC PROBIT. This procedure fits a probit model to the data, with several possible variations. Namely it fits the model

$$p(x) = c + (1 - c)\Phi(\beta_0 + \beta_1 x)$$

where x = concentration or log (concentration), c is the background rate, and β_0, β_1 are unknown parameters to be estimated from the data. $p(x)$ is the response probability corresponding to x . The value of c may be known or unknown. Estimation is done by maximum likelihood, based on binomial theory.

Jennrich and Moore [43] show that for distributions in the exponential family, maximum likelihood calculations can be carried out by means of nonlinear least squares regression calculations. This applies, in particular to models based on the binomial distribution.

Thus dose response curves can be fitted to the data by use of appropriate nonlinear regression programs. Both SAS [12] (PROC NLIN) and BMDP [27] (P3R and PAR) contain nonlinear regression programs that can carry out these calculations. See Jennrich and Moore [43] for a discussion of the theory underlying the relation between maximum likelihood estimation and nonlinear regression in the exponential family. We illustrate the methodology with the use of SAS PROC NLIN. However any nonlinear regression program with capability to carry out iteratively reweighted least squares (i.e. allow weights to be functions of the model parameters) would suffice.

SAS PROC PROBIT also calculates lower and upper confidence bounds on concentration values corresponding to various response curve percentile (after adjusting for background), by use of Fieller's theorem (Finney [11] pp 78-79). We illustrate how these confidence bounds can be calculated, based on the parameter estimates of the fit and their asymptotic variance-covariance matrix.

Before discussing the details of fitting dose response curves by means of nonlinear regression programs. We discuss some of the reasons that one might wish to do this.

1. The data analyst may have a general purpose nonlinear regression program available but no special purpose dose response estimation program. Thus the general tool can be used without modification instead of having to write a special purpose program.

2. A very wide variety of models can be fitted to the data by use of general nonlinear regression. PROC PROBIT is rather limited in the extent of models it will fit. Namely it will fit only a probit model using concentration or log concentration. It will adjust for background effects only by Abbott's correction.

We may wish to fit models other than the probit, e.g. the logit model, or even more complex models that incorporate both the probit and logit models as special cases. Background effects might be modelled as additive concentrations rather than by Abbott's correction. Namely,

$$p(x) = \Phi(\beta_0 + \beta_1 \log(c + c_0))$$

where c_0 , the background effect, represents an alternative way of accounting for background. Such a model, although non-standard, can easily be fitted by nonlinear regression techniques. Also, transformations of concentration levels other than the logarithmic are useful. For example the square root transformation.

3. PROC PROBIT automatically inflates variance and covariance estimates and confidence interval bounds by heterogeneity factors whenever the probit model does not fit the data (as determined by the residual chi square statistic). This is not always what we wish to do. PROC NLIN does not inflate variance estimates by heterogeneity factors.
4. We can calculate and save predicted and residual values and thus easily construct residual plots.

It should be noted that PROC NLIN will not compute confidence intervals on response curve distribution percentiles by use of Fieller's Theorem, as PROC PROBIT does. However we show in the subsequent discussion how we can carry out these calculations fairly easily, using either a hand calculator or a small computer program, once the parameter estimates and their asymptotic variances and covariances have been determined.

We now consider three examples of fitting dose response models to fry mortality data by use of SAS PROC NLIN. We use the Holcombe and Phipps fry mortality data in all the examples. The models fitted are:

$$p(\text{conc}) = c + \bar{c}\Phi(\beta_0 + \beta_1 \text{conc})$$

$$p(\text{conc}) = c + \bar{c} \frac{e^{\beta_0 + \beta_1 \text{conc}}}{1 + e^{\beta_0 + \beta_1 \text{conc}}}$$

$$p(\text{conc}) = \Phi(\beta_0 + \beta_1 \log_{10}(\text{conc} + c))$$

All the models are fitted based on binomial distribution theory after pooling data across tanks within groups. This is the way that PROC PROBIT fits models and is appropriate if there is no tank to tank heterogeneity within groups. In the presence of tank to tank heterogeneity we can first adjust the data by an adjustment factor and then pool across tanks within groups.

The first model fit is a repeat of a model we fitted by PROC PROBIT and serves to verify that we can duplicate the PROC PROBIT fits by nonlinear regression. The second model is a logit model and illustrates that we can fit alternative models with PROC NLIN and that the probit and logit model fits result in very similar inferences.

The third model fit treats the background as an additive concentration rather than adjusting for it by Abbott's correction. This sort of model would be appropriate if the mechanism of response due to background sources is the same as the mechanism of response due to the substance under test.

We now discuss the formulation of fitting dose response curves by means of nonlinear regression techniques.

Suppose that there are I concentration groups (both control and treatment) and that the i -th group contains N_i subjects and has X_i responses. Let $p_i(\theta)$ denote the response probability within the i -th group. Then $X_i \sim \text{Binomial}(N_i, p_i(\theta))$.

The form of $p_i(\theta)$ is specified by the form of the dose response model. For example in model 1, $p_i(\theta) = c + \bar{c}\Phi(\beta_0 + \beta_1 \text{conc}_i)$, where $\theta \equiv (\beta_0, \beta_1, c)$ is the unknown parameter vector, to be estimated by least squares.

Under the model assumptions X_i has mean $\mu_i(\theta)$ and variance $\sigma_i^2(\theta)$, where

$$\mu_i(\theta) = N_i p_i(\theta)$$

$$\sigma_i^2(\theta) = N_i p_i(\theta) (1 - p_i(\theta)).$$

The nonlinear regression procedure optimizes the function

$$Q(\theta) \equiv \sum_{i=1}^I (X_i - \mu_i(\theta))^2 w_i(\theta)$$

where $w_i(\theta) = 1/\sigma_i^2(\theta)$. Jennrich and Moore [43] show that optimizing $Q(\theta)$ by the Gauss-Newton method is equivalent to fitting the dose response

curve by maximum likelihood estimation.

We fit the models to the Holcombe and Phipps compound D data. We first consider, $p_i(\theta) = c + \bar{c}\phi(\beta_0 + \beta_1 \text{conc}_i)$. The results of fitting this model in the standard manner with PROC PROBIT are shown in Figure XIV.11.

We first discuss the NLIN commands needed to produce the output. See the SAS 79 manual [12], pp 317-329, for further details.

The model is

$$\mu_i(\theta) = N_i [c + \bar{c}\phi(\beta_0 + \beta_1 d_i)] \equiv N_i p_i(\theta)$$

where $(\beta_0, \beta_1, c) \equiv \theta$ are to be estimated by weighted least squares. The weights are $w_i \equiv w_i(\theta) = 1/[N_i p_i(\theta)(1 - p_i(\theta))]$. The fitting algorithm also uses the derivatives of the mean value function. These are:

$$\partial \mu / \partial \beta_0 = N_i \bar{c} \phi(\beta_0 + \beta_1 d_i)$$

$$\partial \mu / \partial \beta_1 = N_i \bar{c} d_i \phi(\beta_0 + \beta_1 d_i)$$

$$\partial \mu / \partial c = N_i [1 - \phi(\beta_0 + \beta_1 d_i)]$$

where $\phi(x)$, $\Phi(x)$ represent the standard normal probability density function and cumulative distribution function respectively.

The SAS commands needed to generate this fit are given below

```

1 PROC NLIN BEST=20 METHOD=GAUSS;
2 PARAMETERS BC=-5.0 TO -4.0 BY 0.25
   B1=0.0 TO 0.10 BY 0.025
   C=0.03 TO 0.11 BY 0.02;
3 BOUND 0<=C<=1.0;
   ARG=BC+B1*CONCMEAN;
   ARG=MAX(ARG,-5.0);
   ARG=MIN(ARG,5.0);
   BIGPHI=PROBNORM(ARG);
   SMLPHI=0.3989*EXP(-0.5*ARG**2);
   PROB=C+(1.0-C)*BIGPHI;
4 MODEL DEADFSUM=PROB*FRYSUM;
5 DER.BC=(1.0-C)*SMLPHI*FRYSUM;
   DER.B1=(1.0-C)*CONCMEAN*SMLPHI*FRYSUM;
   DER.C=(1.0-BIGPHI)*FRYSUM;
6 OUTPUT OUT=HOLPHIA PREDICTED=PRDFSM RESIDUAL=RSDFSM;
   WEIGHT=1.0/(FRYSUM*PROB*(1.0-PROB));
7 TITLE2 PROBIT MODEL FIT WITH ABBOTT'S CORRECTION--UNTRANSFORMED CONCENTRATIONS

```

Line 1 instructs NLIN to fit by the Gauss-Newton method. NLIN can also fit models by use of the method of steepest descent or the Marquardt method. The Marquardt method is a compromise between Gauss-Newton and steepest descent. Sometimes near the optimum, steepest descent or Marquardt methods take smaller steps and in different directions than the Gauss-Newton method and so are less prone to overshoot the optimum. Thus one optimization method will sometimes produce convergence when another one does not. The BEST = 20 command instructs NLIN to print out the locations and error sums of squares values that it calculates in the preliminary grid search to determine starting values for the iterative portion of the search.

Line 2 specifies parameter values and/or ranges of parameter values that NLIN should use for a preliminary grid search to arrive at starting values for the iterative phase.

Line 3 specifies bounds on the parameters. If the parameters exceed these bounds at any time during the iterative process they are forced back into the permissible region.

Line 4 contains the model specification. The variable DEADFSUM represents the mortality within the i -th group. FRYSUM is the total number of fish exposed (pooled over tanks within groups) and PROB is the response probability in the i -th group. The SAS program statements between lines 3 and 4 are used in the specification of the model in line 4.

Line 5 contains expressions for the derivatives $\partial\mu/\partial\beta_0$, $\partial\mu/\partial\beta_1$, $\partial\mu/\partial c$ respectively.

Line 6 specifies that the predictions and residuals from the fit be calculated and saved for future use.

Line 7 specifies the weights that are to be used in the weighted least squares fit. Note that these weights are functions of the model parameters (through PROB). They are updated following each iteration.

The output from these commands appears in Figures XV.1 to XV.4. Figure XV.1 contains a listing of the Holcombe and Phipps data. Figure XV.2 contains a summary of the parameter values and residual sums of squares associated with the 20 best points in the preliminary grid search. The point with smallest weighted residual sum of squares is used to start the iterative Gauss-Newton search procedure. The results of the Gauss-Newton iteration are summarized in Figure XV.3. It converges after 8 steps. Figure XV.4 contains statistics based on the model converged to in Figure XV.3. The upper portion of the page contains an analysis of variance table based on weighted sums of squares. The middle portion of the page contains parameter estimates and asymptotic standard errors. The bottom of the page contains the asymptotic correlation matrix among the parameter estimates. We compare these results with those in Figure XIV.11.

Several points need to be remembered in making the comparison.

1. We are fitting the model $p(\theta) = c + \bar{c}\Phi(\beta_0 + \beta_1 \text{conc})$ whereas PROC PROBIT parameterizes the model as $p(\theta) = c + \bar{c}\Phi((\beta_0 - 5) + \beta_1 \text{conc})$. Thus the estimates $\hat{\beta}_1$, \hat{c} in the two fits should agree while the PROC NLIN estimate of $\hat{\beta}_0$ should be 5 smaller than the corresponding PROC PROBIT estimate. Comparison of the estimates shows that this is the case.
2. The residual chi square (0.3361) calculated by SAS PROC PROBIT is the same as the (weighted) residual sum of squares in the PROC NLIN fit. Thus this residual sum of squares provides a test of goodness of fit of the model.
3. The asymptotic variances and covariances calculated by PROC NLIN need to be adjusted before being compared to those calculated by PROC PROBIT. In particular, combining the asymptotic standard errors and the asymptotic correlation matrix obtained by PROC NLIN, we calculate the asymptotic variance covariance matrix.

$$\begin{pmatrix} 0.066928 & -0.000543 & -0.000618 \\ -0.000543 & 0.00000455 & 0.00000468 \\ -0.000618 & 0.00000468 & 0.00002797 \end{pmatrix}$$

This matrix looks nothing like the asymptotic variance covariance matrix that is calculated by SAS PROC PROBIT. The reason for this is as follows. We stated that X_i has mean $\mu_i(\theta) \equiv N_i p_i(\theta)$ and variance $\sigma_i^2(\theta) \equiv N_i p_i(\theta)(1 - p_i(\theta))$. However the weighted least squares fit is carried out assuming that $\text{Var}(X_i) \equiv k\sigma_i^2(\theta)$, where k is a constant to be estimated from the data. Thus the estimates of the variances and covariances given by PROC NLIN assume $\hat{\text{Var}}(X_i) = \hat{k}\sigma_i^2(\hat{\theta})$. Thus all variances and covariances are multiplied by \hat{k} .

How is k estimated? Just as in the case of weighted linear regression, k is estimated by the residual mean square. Namely,

$$\hat{k} = \text{weighted residual mean square} = 0.11203710$$

Our maximum likelihood model, though, tells us that $k = 1$. We thus need to adjust all variances and covariances to this value of k . To do this, we simply divide the above variance covariance matrix by \hat{k} . When this is done we obtain

$$\begin{pmatrix} 0.5974 & -0.00485 & -0.00551 \\ -0.00485 & 0.0000405 & 0.0000418 \\ -0.00551 & 0.0000418 & 0.000250 \end{pmatrix}$$

This matrix is nearly the same as that calculated by PROC PROBIT.

An important purpose for fitting the probit model is to calculate lower confidence bounds on safe concentrations by use of Fieller's theorem (Finney [11], pp 78-79). That is, we wish to calculate a lower bound on the concentration such that $\Phi(\beta_0 + \beta_1 \text{conc}) = L$, where L is some specified response rate. Such lower confidence bounds, at (one sided) confidence level 97.5 percent are a standard part of the PROC PROBIT output. They are given in Figure XIV.12 for the Holcombe and Phipps compound D data with untransformed concentration. We indicate below how to calculate these bounds for any confidence level, based on the output from PROC NLIN. The theory underlying these calculations is sketched in Appendix AXV.

The fitted model is $\hat{p}(\text{conc}) = \hat{c} + \hat{\epsilon}\Phi(\beta_0 + \beta_1 \text{conc})$.

We wish to construct a $1 - \alpha$ level confidence interval on that CONC such that

$$\Phi(\beta_0 + \beta_1 \text{conc}_L) = L$$

where L is specified (e.g. 0.01, 0.05, 0.10 etc.). L represents the response level attributed to toxicant (i.e. over and above background).

The point estimate, conc_L , is

$$\text{conc}_L = (\Phi^{-1}(L) - \hat{\beta}_0)/\hat{\beta}_1 \equiv (f_L - \hat{\beta}_0)/\hat{\beta}_1.$$

Let the asymptotic variance-covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$ be denoted as

$$\begin{pmatrix} g & h \\ h & j \end{pmatrix} \equiv \begin{pmatrix} \widehat{\text{Var}}(\hat{\beta}_0) & \widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) \\ \widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) & \widehat{\text{Var}}(\hat{\beta}_1) \end{pmatrix}$$

A $1 - \alpha$ confidence interval on conc_L is shown in Appendix AXV to be

$$\text{conc}_L \pm \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where $A = \hat{\beta}_1^2 - jz_{\alpha/2}^2$

$$B = 2[\hat{\beta}_1 (\hat{\beta}_0 - f_L) - h z_{\alpha/2}^2]$$

$$C = [(\hat{\beta}_0 - f_L)^2 - g z_{\alpha/2}^2]$$

$z_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution.

The quantities $\hat{\beta}_0$, $\hat{\beta}_1$, g, h, j are obtained as output from NLIN. The results of the calculations are given below.

Holcombe and Phipps -- Compound D -- Untransformed Concentration
Results from calculation of upper and lower 95 percent
confidence bounds on various percentiles of PROBIT fit -- by Fieller's
Theorem

L	$f_L \equiv \Phi^{-1}(L)$	Lower 95% conf.limit	upper 95% conf.limit
0.01	-2.345	22.958	69.385
0.03	-1.88	39.00	78.23
0.05	-1.645	47.07	82.73
0.07	-1.48	52.73	85.91
0.10	-1.28	59.96	89.78
0.15	-1.03	68.06	94.66
0.20	-0.83	74.80	98.62
0.25	-0.68	79.82	101.63
0.50	0	101.69	116.15

These confidence bounds are seen to agree very closely with the bounds calculated by SAS PROC PROBIT and which appear in Figure XIV.12.

The previous dose response model fitted by use of PROC NLIN was a repeat of a model that has also been fitted by PROC PROBIT. Comparisons of the PROC PROBIT and PROC NLIN outputs verified that dose response models can in fact be fitted by nonlinear regression programs and helped to interpret the various features of the PROC NLIN output.

We now consider the fitting two dose response models that cannot be fitted by PROC PROBIT. This of course is the reason for considering the application of PROC NLIN for dose response estimation in the first place. We first consider the logistic model and then look at an alternative to Abbott's correction for accounting for background response.

The logistic model is a commonly used dose response model and gives results very similar to probit fits, at least between the 2nd and 98th percentiles. The logistic c.d.f. is

$$F(x) = \frac{e^x}{1 + e^x} \quad -\infty < x < \infty$$

and is a symmetric unimodal distribution like the normal, but has heavier tails. We fit the dose response model

$$p(\text{conc}) = c + \bar{c}F(\beta_0 + \beta_1 \text{conc})$$

in direct analogy to the probit fit that appears in Figures XIV.11.

The results of the Gauss-Newton iterative process are given in Figure XV.6. The Marquardt algorithm converges whereas the Gauss-Newton algorithm does not because the Marquardt algorithm can take smaller steps and is more flexible in direction. However both algorithms arrive at nearly the same parameter estimates. The summary of the fitted dose response model appears in Figure XV.7. We can compare this fit to the probit fit in Figure XV.4.

We see that both the logit model and the probit model fit the data quite well (residual sums of squares are quite small). The background mortality rate is estimated to be about 0.07 by each model. The asymptotic variance-covariance matrix of the logit fit parameters is estimated to be

$$\begin{aligned} \widehat{\text{Var}} &= \frac{1}{0.093958} \begin{pmatrix} 0.48086 & 0 & 0 \\ 0 & 0.00388 & 0 \\ 0 & 0 & 0.00518 \end{pmatrix} \begin{pmatrix} 1.0000 & -0.988351 & -0.550366 \\ -0.988351 & 1.0000 & 0.520791 \\ -0.55036 & 0.520791 & 1.0000 \end{pmatrix} \\ &= \begin{pmatrix} 0.48086 & 0 & 0 \\ 0 & 0.00388 & 0 \\ 0 & 0 & 0.00518 \end{pmatrix} = \begin{pmatrix} 2.46095 & -0.019626 & -0.01459 \\ -0.019626 & 0.000160 & 0.000111 \\ -0.01459 & 0.000111 & 0.000286 \end{pmatrix} \\ &\equiv \begin{pmatrix} g & h & * \\ h & j & * \\ * & * & * \end{pmatrix} \end{aligned}$$

We now apply Fieller's procedure for calculating lower end upper confidence bounds on distribution percentiles of the dose response fit. We need only modify the calculations done for the probit fit by defining

$$f_L = F^{-1}(L) = \ln \frac{L}{1-L} \equiv \text{logit } (L)$$

and using the appropriate point estimates and variance-covariance matrix.

Holcombe and Phipps -- Compound D -- Untransformed Concentration
Results from calculations of upper and lower 95 percent
confidence bounds on various percentiles of LOGIT fit -- by Fieller's
Theorem

L	$f_L \equiv \ln \frac{L}{1-L}$	lower 95% conf.limit	upper 95% conf.limit	LOGIT point estimate	PROBIT Point estimate
0.01	-4.5951	2.515	66.209	45.31	52.931
0.03	-3.4761	26.859	78.151	61.205	64.231
0.05	-2.9444	38.390	83.863	68.757	69.942
0.07	-2.5867	46.127	87.726	73.838	73.952
0.10	-2.1972	54.526	91.957	79.370	78.812
0.15	-1.7346	64.454	97.031	85.940	84.887
0.20	-1.3863	71.879	100.900	90.888	89.747
0.25	-1.0986	77.967	104.142	94.974	93.392
0.50	0	100.441	117.293	110.578	109.916

The point estimates of the probit and logit fit percentiles are presented side by side for comparison. Except at $L = 0.01$ they are very close and even at $L = 0.01$ they are similar. The situation is a bit different with respect to confidence bounds on the safe concentration. The logit confidence bounds are to be compared with the probit confidence bounds. We see that the upper logit and probit confidence bounds are very similar at each percentile. However the lower confidence bounds for the logit fit are somewhat lower than the lower confidence bounds for the probit fit at the low distribution percentiles.

For $L \geq 0.07$, the lower confidence bounds for the logit and probit fits are rather similar, the lower logit bounds being constantly smaller than the lower probit bounds. For L below 0.05 this phenomenon is accentuated, especially at $L = 0.01$. This is the region in which mortality due to background is the first order effect while toxicant related mortality is secondary. Thus the data and the fitted model reflect primarily the background effects and provide little direct evidence about toxicant related mortality. Since the tails of the logistic distribution are heavier and steeper than the tails of the normal distribution, changes in parameter values perturb percentile estimates in the normal distribution much less than they do in the logistic distribution. Thus

the lower logistic confidence limits become much wider than the corresponding lower normal limits as $L \rightarrow 0$. This phenomenon holds very strongly in this example at $L = 0.01$ and to some extent at $L = 0.03, 0.05$. In this region the data provide little basis to choose between the logit and probit fits. Both models fit the data well and yield very similar point estimates. Thus we learn the following lesson:

THE LOWER CONFIDENCE BOUNDS ON "SAFE" CONCENTRATIONS CORRESPONDING TO LOW PERCENTILES OF THE DOSE RESPONSE DISTRIBUTION MAY BE SENSITIVE TO THE PARTICULAR FORM ASSUMED FOR THE DOSE RESPONSE RELATION, EVEN THOUGH SEVERAL MODELS MAY FIT THE DATA EQUALLY WELL AND PROVIDE SIMILAR POINT ESTIMATES OF PERCENTILES. THE DATA MAY NOT BE SUFFICIENT TO DISTINGUISH BETWEEN THE MODELS.

This phenomenon is observed quite frequently in very low dose extrapolation based on results of carcinogenesis experiments. However in those applications the extrapolation is much more extreme than in fish toxicology applications. However this example illustrates that even in fish toxicology situations the inference about safe dose can be very sensitive to model assumptions, even at the first to the third percentile. The extent of background effects may prevent us from distinguishing among alternative models which fit the data about equally but which yield qualitatively different inferences about safe concentrations corresponding to low distribution percentiles.

To partially circumvent this problem we consider an alternative approach to dose response estimation based on fewer assumptions about the shape of the response distribution. See the following section for a discussion on this nonparametric approach to dose response estimation.

We consider now a third example of fitting dose response models by means of nonlinear regression. This example involves a nonstandard model which provides an alternative to Abbott's correction to account for background response. Abbott's correction is appropriate when the mechanism associated with background effects is independent of the mechanism associated with toxicant effects. For example toxicant mortality may be due to chemical effects whereas background mortality may be due to increased handling of the fish.

However Stephan [42] criticizes the assumption that the control mortality mechanism is totally independent of the toxicant mortality mechanism. He states that stressing the fish during the acclimation or testing periods may make them more susceptible to the toxicant. Thus background effects may act like additions to the toxicant concentrations. Stephan suggests not correcting for control mortality when assessing the effects of various toxicant concentrations.

An alternative way to reflect the dependence between background and toxicant mortality mechanisms is to fit a model which reflects the fact that background may function as an addition to the effective toxicant concentration. Assume that background effects are equivalent to an

addition of c $\mu\text{g/liter}$ in toxicant concentration. The quantity c is a model parameter to be estimated from the data. (Note that the usage of the notation c is completely different in this example than in the previous examples in this section. Here it is being used as a concentration whereas in previous examples the symbol c represented a probability.

Consider the following models based on an assumed normal dose response curve.

$$(1) \quad p(\text{conc}) = \Phi[\beta_0 + \beta_1(\text{conc} + c)]$$

$$(2) \quad p(\text{conc}) = \Phi[\beta_0 + \beta_1(\log_{10}(\text{conc} + c) - 3.0)].$$

These models are to be fitted to the data by maximum likelihood estimation, based on binomial distribution theory. The parameters β_0 , β_1 , c are to be estimated from the data. The first model is over parameterized, in that β_0 and c cannot be separated from one another. Thus to fit model (1) we fit

$$p(\text{conc}) = \Phi[(\beta_0 + \beta_1 c) + \beta_1 \text{conc}] \equiv \Phi(\alpha_0 + \beta_1 \text{conc})$$

using PROC PROBIT with untransformed concentration and no "background" effect included.

The centering constant 3.0 in model (2) is intended to reduce the correlation between $\hat{\beta}_0$, $\hat{\beta}_1$ thereby improving the coverage properties of the fitting algorithms. To fit model (2) we carry out a maximum likelihood analysis using PROC NLIN. The output from this analysis is shown in Figures XV.8, XV.9. The Marquardt algorithm again achieves convergence whereas the Gauss-Newton algorithm does not. Note however that the Gauss-Newton algorithm attains a smaller residual sum of squares due to the difference in weighting. (The distinction between attaining the smallest residual sum of squares and attaining a stationary point corresponds to the difference between minimum chi square estimation and maximum likelihood estimation. This distinction is discussed in Jennrich and Moore [43], page 10, and in the BMDP manual [27]. Both of these methods are asymptotically equivalent. The summary of the Marquardt algorithm fit is presented in Figure XV.9. The residual sum of squares represents a chi square test for goodness of fit of the model. We see that

residual chi square = 59.32 with 3d.f.

Thus the model does not seem to fit the data. We break down this residual chi square into individual cell components to determine whether

the large residual chi square represents consistent lack of fit or the contribution of a single aberrant cell.

We see that the expected frequencies, $N\hat{p}$, $N\hat{q}$, are quite large under the model fit and that there is a systematic discrepancy between model and data. Namely the model underestimates at the lower and upper ends and overestimates p in the middle.

We thus conclude that model (2) is not appropriate for this set of data. However this or similar models may be appropriate for other sets of data. The point is that the use of nonlinear regression techniques to fit dose response curves greatly expands the variety of models that we can fit to the data.

Since model (2) does not fit the data well we do not use it to calculate lower confidence bounds on "safe" dose. However these calculations can easily be made by use of asymptotic maximum likelihood theory.

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION
 EMBRYO AND FRY MORTALITY DATA MERGED WITH AVERAGE CONCENTRATION DATA--THIS IS NOW BASIC DATA
 20:44 TUESDAY, SEPTEMBER 2, 1980

ORS	TRY	EMR SUM	FRY SUM	DEAD SUM	DEAD F SUM	CONC MEAN
1	1	200	100	70	6	0.000
2	3	200	100	70	8	27.215
3	3	200	100	63	8	44.605
4	4	200	100	78	13	72.920
5	0	200	100	67	100	149.150
6						222.950

Figure XV.1 Holcombe and Phipps compound D mortality and concentration data -- pooled

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION
 PROBIT MODEL FIT WITH ABOLIT'S CORRECTION--UNTRANSFORMED CONCENTRATION 20:44 TUE 09/11/80, 5:01 PM 09/11/80

NON-LINEAR LEAST SQUARES GRID SEARCH		DEPENDENT VARIABLE DEAFSUM	
B0	B1	C	WEIGHTED RESIDUAL SS
-5.00	0.050	0.07	27.69964399
-5.00	0.050	0.05	29.38403766
-5.00	0.050	0.09	30.69552631
-5.00	0.050	0.11	36.16508056
-5.00	0.050	0.03	43.52583490
-4.75	0.050	0.07	63.49787243
-4.75	0.050	0.05	65.78624584
-4.75	0.050	0.09	70.41181581
-4.75	0.050	0.03	76.63353606
-4.75	0.050	0.11	77.35752990
-4.00	0.025	0.05	124.66982346
-4.00	0.050	0.07	133.82180978
-4.00	0.050	0.09	137.48705543
-4.00	0.050	0.03	140.38973530
-4.00	0.050	0.11	144.39823039
-4.50	0.025	0.05	147.513799137
-4.50	0.025	0.07	147.75746648
-4.50	0.025	0.09	156.46228056
-4.00	0.025	0.03	170.005530350
-4.25	0.025	0.11	177.92859808

Figure XV.2 Output from SAS proc NLIN applied to Holcombe and Phipps data -- untransformed concentration

FITTING DUSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLIN ADJUSTMENT
 PHUBIT MODEL FIT WITH ABUTTS CORRECTION--UNTRANSFORMED CONCENTRATION 20:44 TUESDAY, SEPTEMBER 2, 1980

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: DEADFSUM METHOD: GAUSS-NEWTON

ITERATION	B0	B1	C	WEIGHTED RESIDUAL SS
0	-5.0000000	0.0500000	0.0700000	27.69964399
1	-4.04611816	0.03484464	0.07138430	5.28213174
2	-4.4229990	0.04013075	0.07128167	0.39137092
3	-4.51674166	0.04102613	0.07172145	0.33634636
4	-4.53388813	0.04114889	0.07176334	0.33612536
5	-4.53388813	0.04115127	0.07176334	0.33611209
6	-4.53388813	0.04115167	0.07176334	0.33611131
7	-4.53388813	0.04115167	0.07176334	0.33611131
8	-4.53388813	0.04115167	0.07176334	0.33611131

NOTE: CONVERGENCE CRITERION MET.

Figure XV.3 Output from SAS PROC NLIN applied to Holcombe and Phipps data -- untransformed concentration

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION
 PROBIT MODEL FIT WITH ABROTT'S CORRECTION--UNTRANSFORMED CONCENTRATION 20:44 TUESDAY, SEPTEMBER 2, 1980 4

NON-LINER LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE DEAFSUM

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	3	64102660.67813536	21367553.55977845
RESIDUAL	6	0.33611139	0.11203710
UNCORRECTED TOTAL		64102661.01424675	
(CORRECTED TOTAL)	5	4498.36044442	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL
B0	-1.52223959	0.25670515	LOWER -5.34656562 UPPER -2.69991057
B1	0.00213252	0.01436492	0.0763842
C	0.00528838	0.05493672	0.0859728

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

	B0	B1	C
B0	1.000000	-0.884592	-0.451540
B1	-0.884592	1.000000	0.414906
C	-0.451540	0.414906	1.000000

Figure XV 4. Output from SAS PROC NLIN applied to Holcombe and Phipps data -- untransformed concentration

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLIN AT REGRESSION 12
 LOGISTIC MODEL FIT WITH ABBOTT'S CORRECTION--UNTRANSFORMED CONCENTRATION
 20:44 TUESDAY, SEPTEMBER 2, 1992

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: DEADFSUM METHOD: GAUSS-NEWTON

ITERATION	B0	B1	C	WEIGHTED RESIDUAL SS
0	-2.0000000	0.0110000	0.0300000	164.29824192
1	-1.9135915	0.01183591	1.734723E-18	138.70015547
2	-2.85612885	0.02423913	1.734723E-18	40.30190465
3	-4.4174540	0.02888243	0.09096944	20.38095723
4	-5.96079151	0.05418146	0.06658499	2.10326736
5	-7.26775002	0.06406083	0.06982378	0.45014653
6	-7.72342051	0.06991161	0.06896286	0.28233120
7	-7.7842001	0.07043157	0.06904970	0.28187440

NOTE: USING THE ABOVE PARAMETERS, THE PARAMETER MODIFICATION VECTOR (DELTA) WAS HALVED 10 TIMES. NO REDUCTION IN THE RESIDUAL SS WAS FOUND (CONVERGENCE ASSUMED).

Figure XV.5 Logit fit to Holcombe and Phipps data using SAS PROC NLIN -- untransformed concentration

FITTING LOGIT RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION 12
 LOGISTIC MODEL FIT WITH ABOTT'S CORRECTION--UNTRANSFORMED CONCENTRATION 20:47 TUESDAY, SEPTEMBER 2, 1980

NON-LINEAR LEAST SQUARES ITERATIVE PHASE
 DEPENDENT VARIABLE: DEADFSUM METHOD: MARQUARDT

ITERATION	B0	B1	C	WEIGHTED RESIDUAL SS
0	-2.00000000	0.01100000	0.03000000	164.29824192
1	-1.51496426	0.01191808	3.469447E-18	137.56798230
2	-1.85944212	0.02438035	3.469447E-18	140.12733427
3	-4.42232597	0.03863354	0.08101047	20.32799412
4	-5.96893004	0.05423865	0.06884728	3.08186413
5	-7.57092834	0.06809747	0.06981510	0.45655573
6	-7.72383965	0.06991643	0.06896285	0.28230946
7	-7.78476594	0.07040193	0.06904992	0.28187490
8	-7.78521738	0.07040485	0.06905248	0.28187385
9	-7.78522188	0.07040491	0.06905251	0.28187383
10	-7.78522632	0.07040491	0.06905254	0.28187382
11	-7.78523069	0.07040494	0.06905257	0.28187381
12	-7.78523502	0.07040497	0.06905260	0.28187380
13	-7.78523930	0.07040500	0.06905263	0.28187380

NOTE: CONVERGENCE CRITERION MET.

Figure XV.6 Logit fit to Holcombe and Phipps data using SAS PROC NLIN -- untransformed concentration

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION 15
 LOGISTIC MODEL FIT WITH ABBOTT'S CORRECTION--UNTRANSFORMED CONCENTRATION
 20:47 TUESDAY, SEPTEMBER 2, 1980

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE DEADFSUM

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	3	291669.41199674	97223.13733225
RESIDUAL	3	0.28181380	0.09395793
UNCORRECTED TOTAL	6	291669.69387054	
(CORRECTED TOTAL)	5	4424.13968364	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL	
			LOWER	UPPER
B0	-7.78223230	0.48085523	-9.31556108	-6.25491753
B1	-0.07050500	0.00388025	0.05805609	0.08275191
C	-0.06905293	0.00517702	0.05257678	0.08552849

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

	B0	B1	C
B0	1.000000	-0.988351	-0.550366
B1	-0.988351	1.000000	0.520791
C	-0.550366	0.520791	1.000000

Figure XV.7 Logit fit to Holcombe and Phipps data using SAS PROC NLIN -- untransformed concentration

```

FITTING DOSE RESPONSE MODELS TO HULCOMHE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION
PROBIT MODEL WITH BACKGROUND TREATED AS ADDITIVE DOSE--LOGARITHMIC CONCENTRATION
20:47 TUESDAY, SEPTEMBER 2, 1980
NON-LINEAR LEAST SQUARES ITERATIVE PHASE
DEPENDENT VARIABLE: DEADFSUM METHOD: MARQUARDT

```

ITERATION	B0	B1	C	WEIGHTED RESIDUAL SS
0	4.0000000	5.0000000	62.0000000	59.91109872
1	3.99701979	5.00832935	61.82606821	59.85623321
2	3.99701979	5.00832935	61.82606821	59.85623321

```

NOTE: CONVERGENCE CRITERION MET.

```

Figure XV.8 Use of SAS PROC NLIN to fit nonstandard dose response model to data

HOLCOMBE AND PHIPPS COMPOUND D FAY MORTALITY DATA

FITTING DOSE RESPONSE MODELS TO HOLCOMBE AND PHIPPS COMPOUND D DATA BY MEANS OF NONLINEAR REGRESSION
 PROBIT MODEL WITH BACKGROUND TREATED AS ADDITIVE DOSE--LOGARITHMIC CONCENTRATION
 20:44 TUESDAY, 5 PT-MPER 2, 1980

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE DEAFESUM

SOURCE	DF	WEIGHTED SS	WEIGHTED MS
REGRESSION	3	1344.22184920	448.07394973
RESIDUAL	6	59.34465332	9.890775553
UNCORRECTED TOTAL	9	1403.56650253	
(CORRECTED TOTAL)	5	1005.24690533	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL
B0	4.13181039	3.03549811	LOWER -5.49483734 UPPER 13.76245511
B1	5.39723127	2.23500790	-17.72808137 28.32274342
C	65.06065330	82.01073320	-195.93850511 328.05981170

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

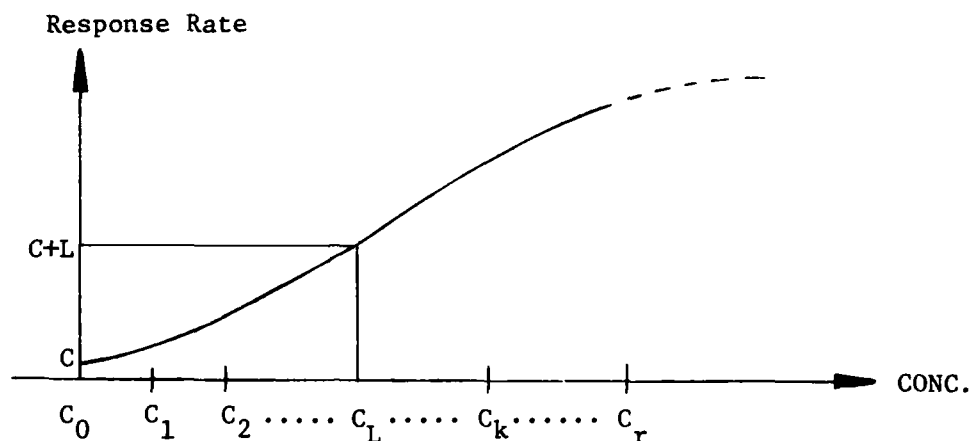
	B0	B1	C
B0	1.000000	0.965543	0.883880
B1	0.965543	1.000000	0.973194
C	0.883880	0.973194	1.000000

Iter	Conc.	N	X	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2}$
1	0.0000	100	6	0.0157	1.568	1.544	12.725
2	27.215	100	8	0.0181	8.876	8.088	0.095
3	44.905	100	8	0.1723	17.234	14.264	5.978
4	72.950	100	13	0.3364	38.698	22.323	19.081
5	128.150	100	79	0.6375	63.746	23.110	10.068
6	222.850	100	100	0.8978	89.784	9.173	11.379
							<u>51.926</u>

Figure XV.9 Output from SAS PROC NLIN to fit nonstandard dose response model to data

XVI. NONPARAMETRIC LOWER CONFIDENCE BOUNDS ON SAFE CONCENTRATIONS

In this section we again consider the estimation of "safe" concentrations based on fitted dose response curves. We wish to estimate a lower confidence bound on that concentration, c_L , for which the response rate is no more than L greater than the control group rate. The value of L is specified by the user. We present this situation pictorially below.



The dose response curve is assumed to be concave upward at the lower percentiles. For a logit or probit fit this would be below the median of the distribution. The solid portion of the illustration represents the region where the dose response curve is concave upward.

The upper bound on the upward concavity region is denoted UCR. UCR is specified by the user.

Let $c_0 \leq c_1 \leq c_2 \leq \dots \leq c_r \leq \text{UCR}$ denote the test concentrations (treatment and control) in the upward concavity region. c_0 , the control group concentration, would often be 0.

The standard method of estimating c_L by means of dose response curves is to assume a specific form for the dose response curve such as probit or logit in concentration or in log concentration and then fit the model by means of maximum likelihood estimation, based on all the data. SAS PROC PROBIT or a nonlinear regression package can be used to fit such models.

The procedure discussed in this section has a number of important differences from these standard parametric dose response models. Among these are

1. Inferences about safe concentrations can sometimes be rather sensitive to the particular form of the dose response curve

assumed. Yet it may not be possible to distinguish among such competing models based on the data at hand. The need for such strong parametric assumptions is alleviated with the procedure in this section.

2. Once a functional form is chosen for the dose response curve, there is still the question of the dose metameter. Different lower bounds may result depending on whether the probit (say) model is chosen with respect to concentration, log (concentration), or some other function of concentration. There is no need to worry about the specific form of the dose metameter with the nonparametric procedure of this section.
3. The parametric dose response models assume a specific functional form for the correction for background responses; Abbott's formula (Finney, [11], pp125-126) is commonly used. However estimates of the low percentiles of the dose response curve can be sensitive to the specific form of background correction used. The procedure in this section does not require the specification of any particular functional form for background response.
4. The standard method of fitting a parametric dose response model is by means of maximum likelihood estimation. The theoretical justification is based on the assumptions of large samples and asymptotic normality. These assumptions may not be entirely satisfied in the case of relatively small sample sizes or of many response group probabilities at or near 0 percent or 100 percent. By contrast, the method discussed in this section is based on exact small sample theory and so is appropriate irrespective of small sample sizes or extreme response rates. We also present an alternative confidence bound calculation which may yield higher lower bounds, however this alternative approach depends on large sample theory and asymptotic normality. Both estimates are routinely calculated by our computer program.
5. The standard parametric probit or logit dose response curve fits utilize the information from all the test concentrations, including those high concentrations at the upper end of the dose response curve, far away from the safe concentration. In fact, these upper concentrations, with high response rates are very instrumental in determination of the slope estimate and associated precision estimate. These high concentrations, thus carry considerable weight, through the specification of the model, in estimating response behavior at the low concentrations. This is not desirable, since the same functional form may not be appropriate throughout the entire range of concentrations. By contrast, the method in this section uses information only from those concentration groups where the dose response curve is concave upward. This is generally in the region below the median of the dose response curve.

One assumption made throughout this section is that the response results can be modelled with the binomial distribution within each tank and that there is no evidence of tank to tank heterogeneity within treatment groups. The responses can then be pooled across tanks within treatment groups and we can assume a single binomial distribution for the pooled responses within each treatment group. This distributional assumption is made in our program.

What do we do if there in fact is evidence of heterogeneity within tanks? There are three approaches to account for this situation. See Section IX for detailed discussion. Briefly these approaches are:

1. Carry out analyses on a per tank basis rather than on a per fish basis. This is the approach that is currently being used by some researchers. However this approach greatly reduces the number of degrees of freedom available for analysis. I feel it is too conservative.
2. Fit distributional models to the data that explicitly account for such tank to tank heterogeneity. Several such models proposed are the beta binomial model (Williams [21]) or the correlated binomial model (Kupper and Haseman [22]). These models generalize the binomial distribution model and can be incorporated into a dose response curve estimation model. The fitting would be by maximum likelihood estimation and the statistical inferences would be based on asymptotic normal distribution theory.
3. We can adjust the data to reflect the within tank correlation. Namely tank to tank heterogeneity reflects itself as variation in response rate from tank to tank within treatment groups. This can also be regarded as correlation of responses within individual tanks. The effect of such correlation is to reduce precisions of estimates as compared to what they would be in a binomial model, since the correlations will usually be positive. This reduced precision can be accounted for in a workmanlike manner by reducing the effective sample size within each tank. Namely suppose we have 4 tanks per group, 25 fry per tank, and responses 1, 3, 8, 7 respectively. The effect of assuming a binomial model would be to pool data across tanks within groups, so that we have 100 fry and 19 responses. Thus $\hat{p} = .19$ and $\sqrt{\text{Var}(\hat{p})} = \sqrt{(.19)(1 - .19)/100} = .039$. However correlation within tanks inflates the variability by a factor h . ($h > 1$). Reduce the assumed sample size within each tank from 25 to $25/h$. Correspondingly reduce the effective number of responses within each tank to $1/h$, $3/h$, $8/h$, $7/h$, for a total of 19/h. Thus $\hat{p} = (19/h)/(100/h) = .19$ still. However $\sqrt{\text{Var}(\hat{p})} = \sqrt{(.19)(1 - .19)/(100/h)} = .039\sqrt{h}$. We then disregard the

tank to tank heterogeneity and utilize the binomial based procedures, such as the computer program discussed in this section.

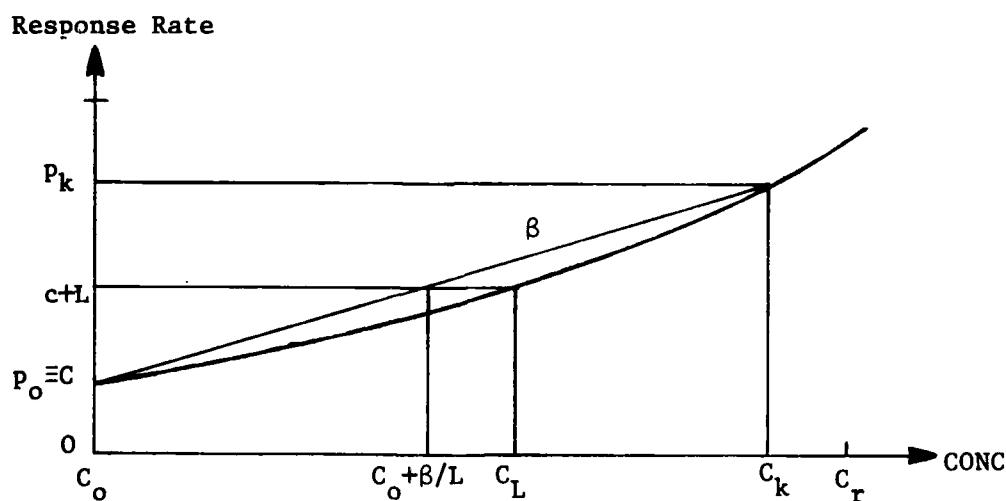
This method of adjustment to effective sample sizes is approximate and somewhat crude, however it has the advantage of simplicity and no special computer programs need be used. Namely, the same methods that are utilized in the absence of tank to tank heterogeneity are used in the presence of such heterogeneity, only with reduced sample sizes. This allows for the use of standard analysis tools in non-standard situations.

Looked at from the perspective of reducing sample sizes to an effective sample size, carrying out analyses on a per tank basis is like reducing the effective sample size in a tank all the way down to 1. I feel that this is going a bit too far.

In particular the methods discussed in this section can be utilized following such adjustments to account for tank to tank heterogeneity. Thus from now on in this section we ignore the question of tank to tank heterogeneity within groups and discuss our procedure, based on binomial distribution theory, as if there were no tank to tank heterogeneity.

We now consider the details of the nonparametric dose response procedure.

Assume that k is such that $c_0 \leq c_L \leq c_k \leq \text{UCR}$. The value of k is specified by the user of the program.



Let p_o, p_k denote the true response rates at c_o, c_k respectively. Draw a chord joining the points $(c_o, p_o), (c_k, p_k)$. Let β denote the slope of this chord. Thus

$$\beta = \frac{p_k - p_o}{c_k - c_o}$$

Upward concavity implies that the chord lies above the dose response curve throughout the region (c_o, c_k) . The concentration at which this chord crosses the value $c + L$ on the response scale is $c_o + L/\beta$. If $c_o = 0$, this concentration is L/β . Thus

$$c_o + L/\beta \leq c_L$$

and so if we can place a lower confidence bound on $c_o + L/\beta$, then this also serves as a nonparametric lower confidence bound on c_L .

Let $\hat{p}_o, \hat{p}_1, \dots, \hat{p}_r$ be the estimated response rates based on binomial theory at concentrations c_o, c_1, \dots, c_r respectively. Then $N_i \hat{p}_i \sim B(N_i, p_i)$ $i = 0, 1, \dots, r$ when N_i is the number of animals on test in the i -th treatment group pooled across tanks and p_i is the true response probability within the i -th treatment group.

Let \underline{p}_o denote a lower confidence bound on p_o , \tilde{p}_k denote an upper confidence bound on p_k . Such exact lower and upper confidence bounds were derived by Clopper and Pearson [46] and are valid for small sample sizes. Expressions for them are contained in a number of sources, including Hollander and Wolfe [47], pages 23, 24. Charts for these confidence intervals are given in a number of places, including Dixon and Massey [13] pp 501-504. Expressions for these confidence bounds are given in Appendix AXVI.1.

An upper confidence bound, β_U , on β is

$$\beta_U = \frac{\tilde{p}_k - \underline{p}_o}{c_k - c_o}$$

Thus a lower confidence bound on c_L is $c_o + L/\beta_U$. This confidence bound is valid in small samples.

The results in the concentration groups $c_{k+1}, c_{k+2}, \dots, c_r$ can be used to improve on the confidence bound discussed above. The details of this procedure, along with a description of an alternative confidence bound, valid in large samples, are discussed in the writeup "A Computer Program to Calculate Nonparametric Lower Confidence Bounds on

Safe Concentrations in Quantal Response Toxicity Tests" by Feder and Sherrill [41]. This writeup also describes in detail the use of a computer program to implement this procedure. This document is included as an appendix to this section. We illustrate the results of this program by example in the remainder of this section and compare the results of the nonparametric estimates of safe concentration with those based on the logit or probit fits.

We first consider the DeFoe compound C fry mortality data.

We have seen from previous sections that there is no evidence of tank to tank heterogeneity within treatment groups.

The various portions of the computer program output are numbered and we discuss them in detail.

As a number of the parameter values used in the program were chosen rather arbitrarily (e.g. UCR) we should regard the output as illustrative of the algorithm's working rather than as a definitive answer in this particular case. We know that the algorithm will give conservative answers. The question is just how conservative the algorithm is.

We know from the preliminary plots and tests of homogeneity that there is no concentration related trend in embryo mortality. Such preliminary analyses are very important to carry out, in order to gain an understanding of the structure of the data. This helps us to interpret the results of the procedures such as the one in this section.

The numbered descriptions below refer to the similarly numbered descriptions in the computer printout for the DeFoe fry mortality data.

1. The title of the output. This title appears at the head of every page.
2. The basic data are presented for each tank within each concentration group (treatment and control). Numbers of fry per tank, numbers survived, and toxicant concentration are given.
3. The number and the proportion of dead fry within each group are given. These values are calculated by pooling across tanks within groups.
4. Basic parameter values for the procedure.

$L \equiv$ response rate, over and above the control rate, at the "safe" concentration.

$k \equiv$ the index of assumed upper bound on the "safe" concentration, c_L .

$$\text{i.e. } c_0 \leq c_L \leq c_k.$$

UCR \equiv upper limit of the concave upward region in the dose response curve.

A number of confidence bound calculations are carried out for differing combinations of (L, k, UCR). In this example UCR is specified as 50. This places it just above the fifth treatment group. ($\hat{p}_5 = .225$, $\hat{p}_6 = 1.000$).

Thus $r = 5$ and $c_r = 48.3074$ in this problem.

L and k are varied.

L = .01, .05, .10
k = 3, 4, 5.

L = .01, k = 3, UCR = 50.0 in the first calculation.

5. Simultaneous confidence interval adjustments are made in this run by means of Bonferroni's inequality with familywise confidence level 0.95. Thus all small sample confidence intervals are calculated at individual confidence level $1 - (.05/4) = 0.9875$.
6. Upper and lower confidence intervals are calculated at each concentration group. These are exact, small sample confidence intervals, calculated as discussed by Clopper and Pearson using the expressions in Appendix AXVI.1.
7. Straight line approximations to the dose response curve are calculated using the combinations of treatment groups shown. The specific method of calculation of the slopes is discussed in the program documentation in Appendix AXVI.2. (Feder and Sherrill [41]). For each combination of concentrations CONC MEAN is the arithmetic average of the concentrations, slope (normal approx) and slope (small sample) are the calculated values of β_U based on either asymptotic theory or exact small sample theory. See the program documentation for details.
8. Lower confidence bounds on c_L are calculated using the minimum of the slopes in paragraph 7 (in this case .0056 for the normal approximation and .0084 for the exact approach). The values given under "calculated safe dose" are $c_1 + L/\beta_U$. These are $.0494 + .01/.0056 = 1.845$ and $.0494 + .01/.0084 = 1.24$ respectively for the normal theory and small sample calculation.

Since we have taken $k = 3$, $c_k \equiv c_3 = 5.9762$ is an upper bound on c_L , by assumption. Thus

$$\hat{c}_L(\text{normal}) = \min(1.85, 5.98) = 1.85$$

$$\hat{c}_L(\text{small sample}) = \min(1.24, 5.98) = 1.24$$

Since the response rates are so extreme in 5 of the 6 groups (i.e. close to 0 or 1) we have small expected frequencies in many of the cells and asymptotic normal theory is suspect here. We will thus confine attention to calculations based on exact small sample theory for the remainder of the section.

9. We modify the parameters defining the procedure. The values k , UCR remain at 3, 50.0 respectively, however L is changed to 0.05. We thus define the "safe" concentration as that which yields a response rate of .05 above control.
10. Proceeding through the same calculations as before we find that the minimum slope (small sample) is .0084. Thus

$$c_1 + L/\beta_U = .0494 + .05/.0084 = 5.98$$

Since $c_k \equiv c_3 = 5.9762$ we estimate

$$\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(5.98, 5.98) = 5.98$$

11. We now alter L to 0.10, leaving k and UCR as before.
12. $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .10/.0084, 5.9762) = \min(11.954, 5.9762) = 5.9762$.

Thus \hat{c}_L is constrained by overly conservative assumption about c_k .

13. We now set $k = 4$ ($c_4 = 14.8125$) and set L back to 0.01
14. We now calculate slopes, but we have fewer to work with. Namely we use c_1, c_4, c_5 in various combinations.
15. $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .01/.0082, 14.8125) = 1.2625$

16. We now change L to 0.05, leaving the other parameters as before.
17. $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .05/.0082, 14.8125) = \min(6.1145, 14.8125) = 6.1145$
18. We change L to 0.10 leaving k , UCR unchanged.
19. $\hat{c}_L = \min(c_1 + L/\beta_U, 14.8125) = \min(.0494 + .10/.0082, 14.8125) = \min(12.1796, 14.8125) = 12.1796$
20. We now change k to 5 and set L back to 0.01.
Thus $c_k \equiv c_5 = 48.3074$.
21. $L = 0.01, k = 5$
 $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .01/.0080, 48.3074) = 1.3045$
22. Change L to 0.05
23. $L = 0.05$.
 $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .05/.0080, 48.3074) = 6.3246$
25. $L = 0.10, k = 5$
 $\hat{c}_L = \min(c_1 + L/\beta_U, c_k) = \min(.0494 + .10/.0080, 48.3074) = 12.5998$

We thus conclude that $k = 3$ is too conservative. Setting $k = 4$ or 5 yields nearly the same lower bound on c_L . In particular for $k = 5$

$L = .01$	$\hat{c}_L = 1.3045$
$L = .05$	$\hat{c}_L = 6.3246$
$L = .10$	$\hat{c}_L = 12.5998$

TABLE 1. SUMMARY OF DATA

- ④ PREPARED BY: [illegible]
- ⑤ [illegible]
- ⑥ [illegible]

⑦ COMPUTED DATA (BASED ON [illegible])

TREATMENT	REFERENCE SOUND	TIME	DATE
1	0.000	0.000	0.000
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.000	0.000	0.000
5	0.000	0.000	0.000
6	0.000	0.000	0.000

⑧ [illegible]

TIME	DATE	TIME	DATE	TIME	DATE
1	0.000	1	0.000	1	0.000
2	0.000	2	0.000	2	0.000
3	0.000	3	0.000	3	0.000
4	0.000	4	0.000	4	0.000
5	0.000	5	0.000	5	0.000
6	0.000	6	0.000	6	0.000

⑨ [illegible]

TIME	DATE	TIME	DATE	TIME	DATE
1	0.000	1	0.000	1	0.000
2	0.000	2	0.000	2	0.000
3	0.000	3	0.000	3	0.000
4	0.000	4	0.000	4	0.000
5	0.000	5	0.000	5	0.000
6	0.000	6	0.000	6	0.000

COLLEGE CURRICULUM VOLUME 2

SECRET

1. 4

[illegible]

12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
84

2

7

1

1
2
3
4
5

—

Abstract

210

121

2

[illegible]

100

[illegible]

5 . 4 2 7

;

● ● ● ● ●

●
●

●
●
●
●
●
●

1. 1941-1942

100

10

100

1

1000

١٠٠

③

—

2000

CONFIDENTIAL

PLANT - CONFIDENTIAL

Figure 1. Schematic representation of the experimental design. The subjects were divided into two groups: the control group and the experimental group. The control group was divided into two subgroups: the control group and the control group. The experimental group was divided into two subgroups: the experimental group and the experimental group.

2018-2019

3150-1-1

1. The first step in the process is to identify the problem or issue that needs to be addressed. This involves gathering information and understanding the context of the problem.

④

104910

7507

4
3
2
1

•

0.1, 0.2, 3

1

①

(R, C, M, L, P, U, X)

(Small sample)

- (SMALL SAMPLE)

2. 11. 1950

六三

44-38861-10

1.26.5

CONFIDENCE INTERVALS FOR MEAN

PARAMETER VALUES

(6) LE 0.0500 KE 4 UCR: 0.0000

NOTE: SIMULTANEOUS CONFIDENCE INTERVALS ARE CALCULATED USING BURN'S CORRECTION

FAMILY CONFIDENCE LEVEL: 0.9500

INDIVIDUAL CONFIDENCE LEVEL(SMALL SAMPLES): 0.9000 INDIVIDUAL CORRECTED FOR LEVEL(NORMAL APPROX): 0.9000

COMPUTED CONFIDENCE INTERVALS ON P (SMALL SAMPLES):

TREATMENT	CONFIDENCE BOUNDS
	LOWER UPPER
1	0.0000 0.0075
2	0.0000 0.0075
3	0.0000 0.0075
4	0.0000 0.0075
5	0.0000 0.0075
6	0.0000 0.0075

PRELIMINARY RESULTS

TREATMENT USED	CONF MEAN	CONF MEAN (APPROX)	SLOPE (SMALL SAMPLES)
1 + 2	0.0000	0.0000	0.0000
1 + 3	0.0000	0.0000	0.0000
1 + 4	0.0000	0.0000	0.0000

SAFE DOSE RESULTS

(7) CALCULATED SAFE DOSE	CALCULATED SAFE DOSE (SMALL SAMPLES)
0.0000	0.0000

STANDARD DEVIATION OF THE DATA

PARAMETER VALUES

(M) σ^2 = 0.0001 K₁ = 1.0000

NOTE: SIMULATED CONFIDENCE INTERVALS ARE BASED ON THE ASSUMPTION OF A NORMAL DISTRIBUTION

FAMILY CONFIDENCE INTERVALS ARE BASED ON

INDIVIDUAL CONFIDENCE INTERVALS (SMALL SAMPLES) = 0.9500 INDIVIDUAL CONFIDENCE INTERVALS (NORMAL APPROX) = 0.9500

COMPUTED CONFIDENCE INTERVALS FOR SMALL SAMPLES

TREATMENT CONFIDENCE INTERVALS

	LOWER	UPPER
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000

PRELIMINARY RESULTS

TREATMENTS USED CONC. MEAN SLOPE (NORMAL APPROX) SLOPE (SMALL SAMPLE)

1 4	7.7710	0.0001	0.0004
1 5	24.5734	0.0076	0.0082
1 6	21.0000	0.0000	0.0000

SAFE VALUE RESULTS

(M) CALCULATED SAFE VALUE (NORMAL APPROX) CALCULATED SAFE VALUE (SMALL SAMPLE)

19.5000	19.5000	19.5000
---------	---------	---------



32

F-A

22
23

3

22

—

•

5

PARAMETER VALUE :

॥

⑧

PROBLEM OF LANGUAGE IN THE HISTORY OF THE HUMAN SCIENCES

FAMILY CONSULTANTS, INC.

1. UNITED STATES OF AMERICA - 1. VOLUME (1970-1971)

CONTINUED ON INSIDE OF COVER OF THIS ISSUE

... TREATMENT ... COMPLETION ...

512

100

10

1961-1962

740700 740700

[illegible]

100

PRELIMINARY RESULTS:

THE-14-PM-000

CONFIDENTIAL - APPENDIX

2001

500

10

SAFE USE & VALUE:

UNCLASSIFIED - AFJ 0052
(INFORMAL - FPRC)

THE NEW YORK PUBLIC LIBRARY

COPIES OF THE
FEDERAL BUREAU OF INVESTIGATION
REPORTS

CHIEF OF POLICE

104722

70655

١٠٠

6. 3246

REPORT - ANALYSIS OF DATA

PARAMETER VALUE

(5)

U: 0.100 K: 0.000 D: 0.000

NOTE: SIMULTANEOUS CONFIDENCE INTERVALS ARE GIVEN ON CRITICAL VALUES AND SIGNIFICANCE PROBABILITY

FAMILY CONFIDENCE LEVEL: 0.9500

INDIVIDUAL CONFIDENCE INTERVAL (SMALL SAMPLE): 0.9500 INDIVIDUAL CONFIDENCE LEVEL (NORMAL APPROX): 0.9500

COMPUTED CONFIDENCE INTERVALS FOR (SMALL SAMPLE):

TREATMENT	CONFIDENCE	LOWER	UPPER
1	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000

PRIMARY RESULTS

TREATMENT USE

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

We now compare the lower confidence bounds on c_L obtained from the nonparametric dose response curve fits to these obtained by more classical probit fits in Section XIV.

Nonparametric	(k = 5, UCR = 50.0)	Asymptotic
	Small Sample	
L = 0.01	$c_L = 1.3045$	1.496
L = 0.05	$c_L = 6.3246$	7.282
L = 0.10	$c_L = 12.5998$	14.515

We fitted a probit model to the data (untransformed concentration) using SAS PROC PROBIT. We corrected for background with Abbott's correction and obtained

	Point estimate	Lower 95% Conf.Bnd.
L = 0.01	3.7725	-157.377
L = 0.05	22.6065	- 84.3307
L = 0.10	32.6468	- 45.9667

The lower bounds based on the probit fit with background are thus useless. We refitted the model with the assumption of no background response. This is therefore a more restrictive model. Results:

	Point estimate	Lower 95% Conf.Bnd.
L = 0.01	1.6567	-17.6457
L = 0.05	21.1299	8.8997
L = 0.10	31.5110	21.2620

Thus at L = .05 and especially at L = 0.10 the nonparametric bounds are more conservative. However they are based on many fewer assumptions.

An attempt was made to fit the probit model to log concentration, as suggested by Finney. The probit program would not converge at all.

We now apply the nonparametric dose response program to the Holcombe and Phipps compound D fry mortality data and compare the estimates of safe concentration with those based on probit and on logit fits. The logic underlying the procedure is indicated in Figure XVI.1

Refer to the computer printout (nonparametric). In this example the control group is at concentration 0 so we do not have to adjust for its affects. However we do have significant background effect.

We see from the listing of the data (① on computer printout following) and mortality rates by group that group 5 has an observed fry mortality rate of 0.79 while group 4 has an observed fry mortality rate

of 0.13. Confidence intervals on these values of p clearly confirm that group 4 is below the median of the dose response curve while group 5 is above the median. Therefore UCR lies somewhere between group 4 and group 5. We have taken it at concentration $100\mu\text{g/liter}$, about midway between the two groups. Thus $c_r = c_4 = 72.9499$.

If we define the "safe" values of L to be below 0.10 (over and above control) we should try $c_k = c_3 = 44.9049$ or $c_k = c_4 = 72.9499$ as upper bounds for c_L . We consider the results of the small sample calculations.

First trying $k = 3$:

$$L = 0.01$$

$$\hat{c}_L = \min(L/\hat{\beta}_U, c_k) = \min(0.01/0.0027, 44.9049) = \underline{3.663}$$

$$L = 0.05$$

$$\hat{c}_L = \min(L/\hat{\beta}_U, c_k) = \min(0.05/0.0027, 44.9049) = \underline{18.3315}$$

$$L = 0.10$$

$$\hat{c}_L = \min(L/\hat{\beta}_U, c_k) = \min(0.10/0.0027, 44.9049) = \underline{36.6629}$$

If we next try $k = 4$, we have less of an adjustment for simultaneity and so we get slightly shorter intervals in this case. Namely

$k = 4$:

$$L = 0.01$$

$$\hat{c}_L = 3.8453$$

$$L = 0.05$$

$$\hat{c}_L = 19.2264$$

$$L = 0.10$$

$$\hat{c}_L = 38.4529$$

Thus $k = 3$ and $k = 4$ yield essentially the same results for all practical purposes.

Let's now compare these results with those obtained by fitting probit and logit models to the data.

Probit models were fitted to the data vs CONC (untransformed) and \log_{10} (CONC). Both probit fits have nonsignificant residual chi square of about the same magnitude and so are judged to fit the data about equally well. The following 95 percent lower confidence bounds on safe concentration were obtained from these fits.

PROBIT FITS

<u>L</u>	<u>UNTNSFMD CONC</u>	<u>LOG10 CONC</u>
0.01	22.403	45.764
0.05	46.147	57.470
0.10	58.715	64.836

We see that the nonparametric fit is much more conservative than the probit fits since it is based on low dose linearity. Note that especially at the low percentiles the probit inference is quite sensitive to whether untransformed concentration or log concentration are used.

We also fitted a logistic dose response to the data using the SAS nonlinear regression program, PROC NLIN. Only untransformed concentrations were used with the logistic fit. Background response was adjusted for the Abbott's correction, just as we did with the probit fit. The following summarizes the results of the logistic fit and compare the inferences with those based on the probit fits.

Chi square for goodness of fit 0.282 with 3d.f.

Thus there is no evidence of lack of fit of the model to the data and the logit and probit models fit about equally well.

Point estimates of percentiles (after adjusting for background)

L	Probit(Untransformed CONC)	Probit(log CONC)	Logit(untransformed CONC)
0.01	53.385	60.974	45.31
0.05	69.946	71.826	68.757
0.10	78.774	78.380	79.370

Thus, except for $L = 0.01$ where there is a bit of disparity among the models (although not of practical concern) the three models yield essentially the same percentile estimates.

We now consider lower confidence bounds on these same percentiles based on the three parametric fits and on the nonparametric fit.

HOLPHI6 - ANALYSIS OF FRY DATA

TRT	REP	NFMR	NHATCH	NNORMH	NFRY	NALIVE	NNURML	TOXICANT CONC
0								
1	1	50.	33.	33.	25.	23.	23.	0.0
1	2	50.	29.	23.	25.	23.	23.	0.0
1	3	50.	38.	36.	25.	24.	24.	0.0
1	4	50.	30.	30.	25.	24.	24.	0.0
TREATMENT TOTALS:		NDEADMB=	70.	PDEADMB=	0.350			
		NDEADFRY=	6.	PDEADFRY=	0.040			
2	1	50.	21.	31.	25.	22.	22.	27.2149
2	2	50.	36.	35.	25.	24.	24.	27.2149
2	3	50.	34.	33.	25.	23.	23.	27.2149
2	4	50.	29.	28.	25.	23.	23.	27.2149
TREATMENT TOTALS:		NDEADMB=	70.	PDEADMB=	0.350			
		NDEADFRY=	8.	PDEADFRY=	0.080			
3	1	50.	35.	35.	25.	23.	23.	44.9049
3	2	50.	38.	38.	25.	25.	25.	44.9049
3	3	50.	26.	25.	25.	20.	20.	44.9049
3	4	50.	38.	36.	25.	24.	24.	44.9049
TREATMENT TOTALS:		NDEADMB=	63.	PDEADMB=	0.315			
		NDEADFRY=	8.	PDEADFRY=	0.080			
4	1	50.	30.	30.	25.	21.	21.	72.9499
4	2	50.	29.	29.	25.	20.	20.	72.9499
4	3	50.	27.	25.	25.	21.	21.	72.9499
4	4	50.	36.	35.	25.	25.	25.	72.9499
TREATMENT TOTALS:		NDEADMB=	78.	PDEADMB=	0.390			
		NDEADFRY=	13.	PDEADFRY=	0.130			
5	1	50.	32.	31.	25.	2.	2.	128.1500
5	2	50.	31.	27.	25.	4.	4.	128.1500
5	3	50.	36.	35.	25.	6.	9.	128.1500
5	4	50.	34.	33.	25.	6.	6.	128.1500
TREATMENT TOTALS:		NDEADMB=	67.	PDEADMB=	0.335			
		NDEADFRY=	79.	PDEADFRY=	0.750			
6	1	50.	36.	0.	25.	0.	0.	222.8500
6	2	50.	34.	0.	25.	0.	0.	222.8500
6	3	50.	25.	0.	25.	0.	0.	222.8500
6	4	50.	29.	0.	25.	0.	0.	222.8500
TREATMENT TOTALS:		NDEADLMA=	76.	PDEADLMA=	0.210			
		NDEADFRY=	10.	PDEADFRY=	1.000			

HOLPHIA - ANALYSIS OF FRY DATA

PARAMETER VALUES:

L: 0.0100 K: 3 UCR: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVEL IS CALCULATED USING BUNFERRMAN INEQUALITY
CONFIDENCE LEVEL: 0.9833

2. COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	
	LOWER	UPPER
1	0.0203	0.1321
2	0.0326	0.1591
3	0.0326	0.1581
4	0.0672	0.2193
5	0.6391	0.8704
6	0.9599	1.0000

PRELIMINARY RESULTS:

TREATMENTS USED	CONC. MEAN	SLOPE (NORMAL APPROX)	SLOPE (SMALL SAMPLE)
1 3	22.4525	0.0022	0.0031
1 4	36.4750	0.0022	0.0027
1 3 4	39.2849	0.0021	0.0023

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CHOSEN SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSEN SAFE DOSE (SMALL SAMPLE)
4.8295	4.8295	3.6663	3.6663

HOLPHIA - ANALYSIS OF FRY DATA

PARAMETER VALUES:

L: 0.0500 K: 3 UCR: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVEL IS CALCULATED USING BONFERRONI INEQUALITY

CONFIDENCE LEVEL: 0.9833

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	
	LOWER	UPPER
1	0.0203	0.1321
2	0.0326	0.1581
3	0.0376	0.1581
4	0.0672	0.2193
5	0.6891	0.8704
6	0.9599	1.0090

PRELIMINARY RESULTS:

TREATMENTS USED	CONC MEAN	SLOPE (NORMAL APPROX)	SLOPE (SMALL SAMPLE)
1 3	22.4525	0.0022	0.0031
1 4	36.4750	0.0022	0.0027
1 3 4	39.2849	0.0021	0.0028

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CHOSEN SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSEN SAFE DOSE (SMALL SAMPLE)
26.1473	24.1473	18.3315	18.3315

HOLPHI6 - ANALYTICAL KEY DATA

PARAMETER VALUES:

L: 0.1000 K: 2 UCR: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVEL IS CALCULATED USING BONFERRONI INEQUALITY
CONFIDENCE LEVEL: 0.9833

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	
	LOWER	UPPER
1	0.0203	0.1321
2	0.0326	0.1581
3	0.0326	0.1581
4	0.0672	0.2193
5	0.4891	0.8704
6	0.9599	1.0000

PRELIMINARY RESULTS:

TREATMENTS USED	CONC. MEAN	SLOPE(NORMAL APPROX)	SLOPE(SMALL SAMPLE)
1 3	22.4525	0.0022	0.0031
1 4	36.4750	0.0022	0.0027
1 3 4	39.2849	0.0021	0.0028

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CHOSN SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSN SAFE DOSE (SMALL SAMPLE)
48.2947	44.9049	36.6629	36.6629

HOLPH16 - ANALYSIS OF FRY DATA

PARAMETER VALUES:

L: 0.0100 K: 4 UCR: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVELS IS CALCULATED USING EDUFERRONI INEQUALITY

CONFIDENCE LEVEL: 0.9750

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	
	LOWER	UPPER
1	0.0223	0.1260
2	0.0352	0.1516
3	0.0352	0.1516
4	0.0711	0.2120
5	0.6971	0.8651
6	0.9638	1.0000

PRELIMINARY RESULTS:

TREATMENTS USED	CONC. MEAN	SLOPE(NORMAL APPROX)	SLOPE(SMALL SAMPLE)
1 4	36.4750	0.0021	0.0026

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CHOSN SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSN SAFE DOSE (SMALL SAMPLE)
4.8410	4.8410	3.8453	3.8453

HOLPH16 - ANALYSIS OF FRY DATA

PARAMETER VALUES:

L: 0.0500 K: 4 USE: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVEL IS CALCULATED USING BONFERRONI INEQUALITY

CONFIDENCE LEVEL: 0.9750

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT CONFIDENCE BOUNDS

LOWER UPPER

1 0.0223 0.1260

2 0.0352 0.1516

3 0.0352 0.1516

4 0.0711 0.2120

5 0.0571 0.0851

6 0.0638 1.0000

PRELIMINARY RESULTS:

TREATMENTS USED CONC. MEAN SLOPE (NORMAL APPROX.) SLOPE (SMALL SAMPLE)

1 4 36.4750 0.0021 0.0024

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE CHOSEN SAFE DOSE CALCULATED SAFE DOSE CHOSEN SAFE DOSE
(NORMAL APPROX.) (NORMAL APPROX.) (SMALL SAMPLE) (SMALL SAMPLE)

24.2049 24.2049 19.2264 19.2264

HOLCOMBE AND PHIPPS COMPOUND D FRY MORTALITY DATA

HOLPH16 - ANALYSIS OF FRY DATA

PARAMETER VALUES:

L: 0.1000 K: 4 UCR: 100.0000

NOTE: SIMULTANEOUS CONFIDENCE LEVEL IS CALCULATED USING BONFERRONI INEQUALITY

CONFIDENCE LEVEL: 0.9750

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT CONFIDENCE BOUNDS

LOWER UPPER

1 0.0223 0.1260

2 0.0352 0.1516

3 0.0352 0.1516

4 0.0711 0.2120

5 0.0971 0.2651

6 0.9638 1.0000

PRELIMINARY RESULTS:

TREATMENTS USED CONC. MEAN SLOPE(NORMAL APPROX) SLOPE(SMALL SAMPLE)

1 4 36.4750 0.0021 0.0026

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE CHOSEN SAFE DOSE CALCULATED SAFE DOSE CHOSEN SAFE DOSE
(NORMAL APPROX) (NORMAL APPROX) (SMALL SAMPLE) (SMALL SAMPLE)

48.4098 48.4098 38.4529 38.4529

Output from special purpose program to calculate lower bounds on safe concentration based on nonparametric dose response curve fit

Lower confidence bounds*on percentiles (after adjusting for background)

L	Probit(Untransformed CONC)	Probit (log CONC)	Logit(Untransformed CONC)	Non- parametric
0.01	22.403	45.764	2.515	3.845
0.05	46.147	57.470	38.390	19.226
0.10	58.715	64.836	54.526	38.453

*Confidence intervals for parametric fits are based on 95% two sided non-simultaneous procedures using Fieller's theorem. Nonparametric confidence intervals are based on simultaneous procedure with familywise confidence level of 95%.

Comparing the lower confidence bounds on safe concentration we see the following relations.

1. The nonparametric procedure is more conservative than the parametric procedures, especially at the very low percentiles. This is because the nonparametric procedure is built on the assumption of linear approximations in the low concentration region, while the logit and probit fits approach their limiting values through exponential decay. Which estimate is more appropriate would need to be a matter of biological judgement. Since responses in the region of concentration for which $L \leq 0.05$ are dominated by background response, the data themselves do not provide much empirical evidence.
2. The results at $L = 0.01$ are surprisingly inconsistent across model fits. There is at least an order of magnitude difference in confidence bounds based on the logit and probit fits, despite the fact that the point estimates are in good agreement. There is even a factor of two difference between the bounds based on the two probit fits, despite good agreement of the point estimates. Thus the inferences at low percentiles are very sensitive to the model assumed. The nonparametric procedure, while conservative, is based on many fewer assumptions.

In this example the probit model fitted the data quite nicely and yielded more liberal confidence bounds on c_L than did the nonparametric procedure. It is our experience that this is not always the case. In some situations the probit or logit model does not fit the data well. In other situations Fieller's method may yield lower confidence bounds on c_L which are negative! We saw this in the case of the DeFoe data. In such cases the nonparametric procedure can provide more liberal

bounds than the parametric procedures. We will see this in the following example.

We now place lower confidence bounds on the safe concentration for the Benoit compound A fry mortality data. We first consider the non-parametric procedure and then compare results with bounds based on the probit model. Figures XVI.2, XVI.3 contain plots of proportions of dead fry vs untransformed concentration and vs $\log_{10}(\text{concentration})$ respectively.

Based on the appearance of these plots, a probit dose response model does not seem to hold very well, especially with respect to untransformed concentration. Furthermore, there is some question about homogeneity of response rates within tanks at the highest concentration.

We first consider the nonparametric procedure. We see that the average concentration at the control group is 0.0809 and there was no observed fry mortality there. Based on the plots of proportion dead fry vs concentration and based on fry mortality proportions printed out by the program, we set UCR, the upper bound on the upward concavity region, to be somewhere above the 5th treatment group. In particular we set $UCR = 15.0$. Then $c_r \equiv c_5 = 13.3182$.

Note that for the purpose of illustration we are assuming that there is no tank to tank heterogeneity within groups. This assumption needs to be checked and appropriate modifications made, if necessary.

Because of the many sample proportions close to 0 the asymptotic normality assumption is questionable and so we use the small sample confidence bounds.

Based on the observed proportions of dead fry in the various groups, if L is less than or equal to 0.10 it makes sense to choose c_k , the upper bound on safe concentrations to be c_4 or c_5 . For definiteness we choose c_4 here. Thus we have

$$UCR = 15.0 \quad k = 4$$

$$\begin{aligned} L = 0.01 \quad \hat{c}_L &= \min(c_1 + L/\hat{\beta}_U, c_k) = \min(0.0809 + 0.01/0.0423, 6.6020) \\ &= \min(0.0809 + 0.2364, 6.6020) = \underline{0.3137} \end{aligned}$$

$$\begin{aligned} L = 0.05 \quad \hat{c}_L &= \min(c_1 + L/\hat{\beta}_U, c_k) = \min(0.0809 + 1.1819, 6.6020) \\ &= \underline{1.2628} \end{aligned}$$

$$\begin{aligned} L = 0.10 \quad \hat{c}_L &= \min(c_1 + L/\hat{\beta}_U, c_k) = \min(0.0809 + 2.3638, 6.6020) \\ &= \underline{2.4447} \end{aligned}$$

We now compare these results with those based on the probit fit, using SAS PROC PROBIT.

First models including background variation were assumed. Namely

$$p(\text{CONC}) = c + \bar{c}\Phi((\beta_0 - 5) + \beta_1 \text{CONC})$$

$$p(\text{CONC}) = c + \bar{c}\Phi((\beta_0 - 5) + \beta_1 \log_{10}(\text{CONC})).$$

The maximum likelihood fitting algorithm was unable to converge with either of these three parameter models!

Next the background rate was specified to be 0 and two parameter probit models were fitted to the data. The plot of proportion dead fry vs CONC suggests that the probit model does not fit the untransformed concentration and in fact the model converged to shows substantial lack of fit to the data. We therefore consider the probit fit in log concentration. This fit is better, but still exhibits marginal statistical evidence of lack of fit (Residual chi square = 8.00 with 4d.f., which is significant at $\alpha = 0.09$). Based on this fit, the 95% lower confidence bounds on response distribution percentiles (using Fieller's theorem and adjusted by Finney's heterogeneity factor) are:

L	0.01	$\hat{c}_L = 0.082$
	0.05	$\hat{c}_L = 0.523$
	0.10	$\hat{c}_L = 1.362$

Thus in this example these bounds are lower than those based on the nonparametric procedure. They are also based on a much more restrictive model.

The nonparametric procedure seems quite superior in this example.

RESULTS - ANALYSIS OF FRY DATA

TRE REP * TESTED * SURVIVED * TREATMENT CONC

1	1	15.	15.	0.0000
1	2	15.	15.	0.0000
1	4	15.	15.	0.0000

TREATMENT TOTALS: TOTAL #DEAD: 0.0
PROPORTION DEAD: 0.0

2	1	15.	15.	1.6992
2	2	15.	14.	1.6992
2	4	15.	15.	1.6992
2	4	15.	15.	1.6992

TREATMENT TOTALS: TOTAL #DEAD: 1.000
PROPORTION DEAD: 0.017

3	1	15.	14.	3.1892
3	2	15.	14.	3.1892
3	4	15.	15.	3.1892
3	4	15.	15.	3.1892

TREATMENT TOTALS: TOTAL #DEAD: 0.000
PROPORTION DEAD: 0.033

4	1	15.	12.	6.6020
4	2	15.	11.	6.6020
4	4	15.	14.	6.6020
4	4	15.	14.	6.6020

TREATMENT TOTALS: TOTAL #DEAD: 9.000
PROPORTION DEAD: 0.150

5	1	15.	9.	13.3182
5	2	15.	9.	13.3182
5	4	15.	9.	13.3182
5	4	15.	8.	13.3182

TREATMENT TOTALS: TOTAL #DEAD: 25.000
PROPORTION DEAD: 0.467

6	1	15.	8.	26.6181
6	2	15.	13.	26.6181
6	4	15.	12.	26.6181
6	4	15.	10.	26.6181

TREATMENT TOTALS: TOTAL #DEAD: 27.000
PROPORTION DEAD: 0.480

INDIVIDUAL CONFIDENCE LEVEL DATA

PARAMETER VALUES:

L: 0.0100 K: 3 UCP: 15.0000

NOTE: SIMULTANEOUS CONFIDENCE INTERVALS FOR 0.000010 USING BONFERRONI INEQUALITY

FAMILY CONFIDENCE LEVEL: 0.9990

INDIVIDUAL CONFIDENCE LEVEL (SMALL SAMPLE): 0.9975 INDIVIDUAL CONFIDENCE LEVEL (NORMAL APPROX): 0.9979

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS
	LOWER UPPER
1	0.0 0.0194
2	0.0002 0.1017
3	0.0028 0.1286
4	0.0631 0.2828
5	0.3232 0.6175
6	0.3049 0.6015

PRELIMINARY RESULTS:

TREATMENTS USED	LOG MEAN	LOGP (NORMAL APPROX)	LOGP (SMALL SAMPLE)
1 2	1.0251	0.0290	0.0414
1 4	3.3415	0.0403	0.0434
1 5	6.5995	0.0472	0.0466
1 3 4	3.2507	0.0408	0.0440
1 3 5	5.8294	0.0500	0.0501
1 4 5	6.6870	0.0473	0.0459
1 2 4 5	5.7976	0.0491	0.0503

SAFE DOSE RESULTS:

LOG MEAN	LOGP (NORMAL APPROX)	LOGP (SMALL SAMPLE)	LOGP (SMALL SAMPLE)
0.4259	0.4259	0.3225	0.3225

RESULTS - ANALYSIS OF DATA

PARAMETER VALUES

LS: 0.000 KS: 0 UCR: 12.000

NOTE: SIMULATIONS CONSIDERED INDEPENDENT AND UNCORRELATED WITH EACH OTHER INEQUALITY

FAMILY CONFIDENCE LEVEL: 0.9500

INDIVIDUAL CONFIDENCE LEVEL (SMALL SAMPLE): 0.9075 INDIVIDUAL CONFIDENCE LEVEL (NORMAL APPROX): 0.9000

COMPUTED CONFIDENCE BOUNDS ON τ (SMALL SAMPLE):

TREATMENT CONFIDENCE BOUNDS

	LOWER	UPPER
1	0.0	0.0704
2	0.0002	0.1017
3	0.0025	0.1286
4	0.0031	0.2025
5	0.0202	0.0375
6	0.0044	0.0015

PRELIMINARY RESULTS:

TREATMENT GROUP LONG MEAN SLOPE (NORMAL APPROX) SLOPE (SMALL SAMPLE)

1 3	1.0551	0.0290	0.0414
1 4	5.3415	0.0453	0.0434
1 5	6.6955	0.0472	0.0466
1 6	1.2057	0.0405	0.0440
1 3 5	5.5294	0.0500	0.0501
1 4 5	6.6270	0.0473	0.0468
1 3 4 5	5.7970	0.0491	0.0503

SAFE DISE RESULTS:

COMPUTED DISE RATE (NORMAL APPROX) CHOSEN SAFE DISE RATE (NORMAL APPROX) CALCULATED DISE RATE (SMALL SAMPLE) CHOSEN SAFE DISE RATE (SMALL SAMPLE)

1.0050 1.0050 1.0291 1.2891

2447

RESULTS - ANALYSIS OF TEST DATA

PARAMETER VALUES:

LS: 0.0000 0.0000 0.0000 0.0000 0.0000

NOTE: SIMULTANEOUS CONFIDENCE INTERVALS ARE BASED ON A JOINTLY DISTRIBUTED NORMAL INEQUALITY

FAMILY CONFIDENCE LEVEL: 0.9500

INDIVIDUAL CONFIDENCE LEVEL (SMALL SAMPLE): 0.9500 INDIVIDUAL CONFIDENCE LEVEL (NORMAL APPROX): 0.9500

COMPUTED CONFIDENCE BOUNDS ON F (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	LOWER	UPPER
1	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000

PRELIMINARY RESULTS:

TREATMENTS USED	CONF MEAN	SLOPE (NORMAL APPROX)	SLOPE (SMALL SAMPLE)
1 4	0.0000	0.0000	0.0000
1 5	0.0000	0.0000	0.0000
1 4 5	0.0000	0.0000	0.0000

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)
0.0000	0.0000

Wabstus, I. J. v. L.

.....

NOTE: 1. THE FOLLOWING INFORMATION IS FOR THE USE OF THE READER ONLY.

FM 44-17 (MFR) 10-1-78 L-V-L: J64551

95% NORMAL CONFIDENCE LEVEL (NORMAL APPROX): 0.9833

COMPUTED FOR MODEL RUNS ON 11/11/77

TREATMENT	CONFIDENCE	LOWER	UPPER
1	0.000	0.0000	0.0000
2	0.003	0.0000	0.0060
3	0.003	0.0000	0.0060
4	0.062	0.0000	0.1250
5	0.268	0.0000	0.5355
6	0.215	0.0000	0.4300

MR. LIP. VARY K. JULL, :

PLATE 4. INTS. OF 10	CLINC. MEAN	SLOPE (NORMAL APPRX)	SLOPE (SMALL SAMPLE)
1 -	3.3410	0.0380	0.0423
1 -	2.8995	0.0456	0.0461
1 - 5	6.670	0.0497	0.0463

State - 1960-1961

DATE	DESCRIPTION	AMOUNT	BALANCE
1954-1-1	OPENING BALANCE	100.00	100.00
1954-1-15	PAYROLL	25.00	75.00
1954-1-30	RENT	15.00	60.00
1954-2-15	SALES	50.00	110.00
1954-2-28	PAYROLL	25.00	85.00
1954-3-15	RENT	15.00	70.00
1954-3-31	CLOSING BALANCE	70.00	70.00

ANALYSIS OF DATA

PARAMETER VALUES

LE: 0.000000 K: 5 DUSE: 1.000000

NOTE: SIMULATIONS CONFIDENCE INTERVALS ARE CALCULATED USING NONPARAMETRIC INEQUALITY

FAMILY CONFIDENCE BY LE: 0.95000
INDIVIDUAL CONFIDENCE LEVEL(SMALL SAMPLE): 0.9750 INDIVIDUAL CONFIDENCE LEVEL(NORMAL APPROX): 0.9500

COMPUTED CONFIDENCE BOUNDS ON P (SMALL SAMPLE):

TREATMENT	CONFIDENCE BOUNDS	
	LOWER	UPPER
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0710	0.2657
5	0.3367	0.6000
6	0.3210	0.5839

PRELIMINARY RESULTS:

TREATMENTS USED	CONC MEAN	SLOPE(NORMAL APPROX)	SLOPE(SMALL SAMPLE)
1 5	0.6945	0.0423	0.0483

SAFE DOSE RESULTS:

CALCULATED SAFE DOSE (NORMAL APPROX)	CHOSEN SAFE DOSE (NORMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSEN SAFE DOSE (SMALL SAMPLE)
0.2121	0.2121	0.2015	0.2015

二、「**三民主義**」

[illegible]

NOTE: 5. DIFFERENCES BETWEEN THE CALCULATED USING BOUTER'S INEQUALITY

١٠٥

[illegible][illegible]

COMPUTED: C:\WINDOWS\SYSTEM32\cmd.exe

TREATMENT : CONFIDENTIAL

U.S. DEPT. OF JUSTICE

255

1. 6. 64 2. 3. 64

—

140 00155

15370 2160

0.65E-07

DECLASSIFIED

THE 44-1555

1

• 225 •

6-14-22

254)

SATC L-51-11113

REGULATORY AFFAIRS

—

2000

11-11-77

15. 77

CALCULATED SAFE DISE CHUSAN SAT. USE

Calculus 5A-6

Calculus 5416

CHURCHMAN, SAUL, D.D.

PAKISTAN VALU. 33

1000
1000
1000
1000
1000

NOTE: SIMULTANEOUS CONFIDENCE INTERVALS ARE CALCULATED USING BONFERRONI ADJUSTMENT (P=0.005)

FAMILY CONFIDENCE LEVEL: 0.9500
INDIVIDUAL CONFIDENCE LEVEL(SAMPLE): 0.9500
INDIVIDUAL CONFIDENCE LEVEL(NORMAL APPROX): 0.9500

COMPUTED CONTINGENCY FINDING 11.1 (SMALL SAMPLE):

TREATMENT	CONFIDENCE LOWER	CONFIDENCE UPPER
1	0.000	0.0000
2	0.000	0.0000
3	0.000	0.0000
4	0.000	0.0000
5	0.000	0.0000
6	0.000	0.0000

PRELIMINARY RESULTS:

TESTING UNIT USED	CURC MEAN	SLOPE (NORMAL APPROX)	SLOPE (SMALL SAMPLE)
5	6.8995	0.1439	0.0453

SAFE CLUB RESULTS:

CALCULATED SAF. DOSE (MINIMAL EFFECT)	CHOSEN SAF. DOSE (THERMAL APPROX)	CALCULATED SAFE DOSE (SMALL SAMPLE)	CHOSEN SAFE DOSE (SMALL SAMPLE)
2.0027	2.0027	2.0270	2.0270

Benoit-Compound A

BENOIT1 - PROP DEAD FRY VS AVG CONC OVER TRT GROUP

12:19 WEDNESDAY, SEPTEMBER 10, 1980

PLOT OF PRPUEADF*CONCMEAN LEGEND: A = 1 OBS, B = 2 OBS, ETC.

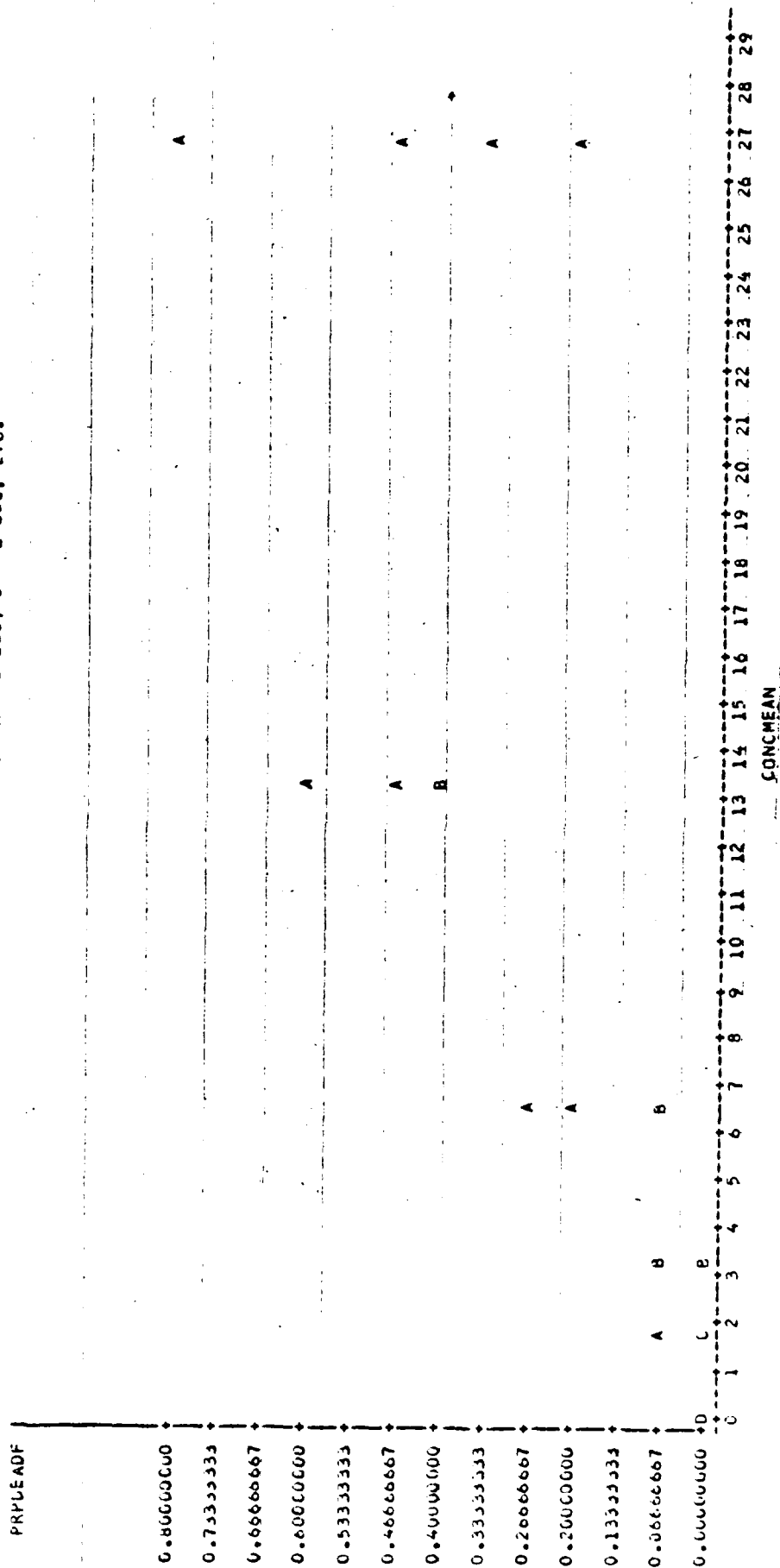


Figure XVI.2

Besit - Compound A

86NO111 - PROP DEAD FRY VS AVG LOG CONC OVER TRT GROUP

12:19 WEDNESDAY, SEPTEMBER 10, 1980

PLUT OF PRPDEADF*LOGCONC

LEGEND: A = 1 OBS, B = 2 OBS, ETC.

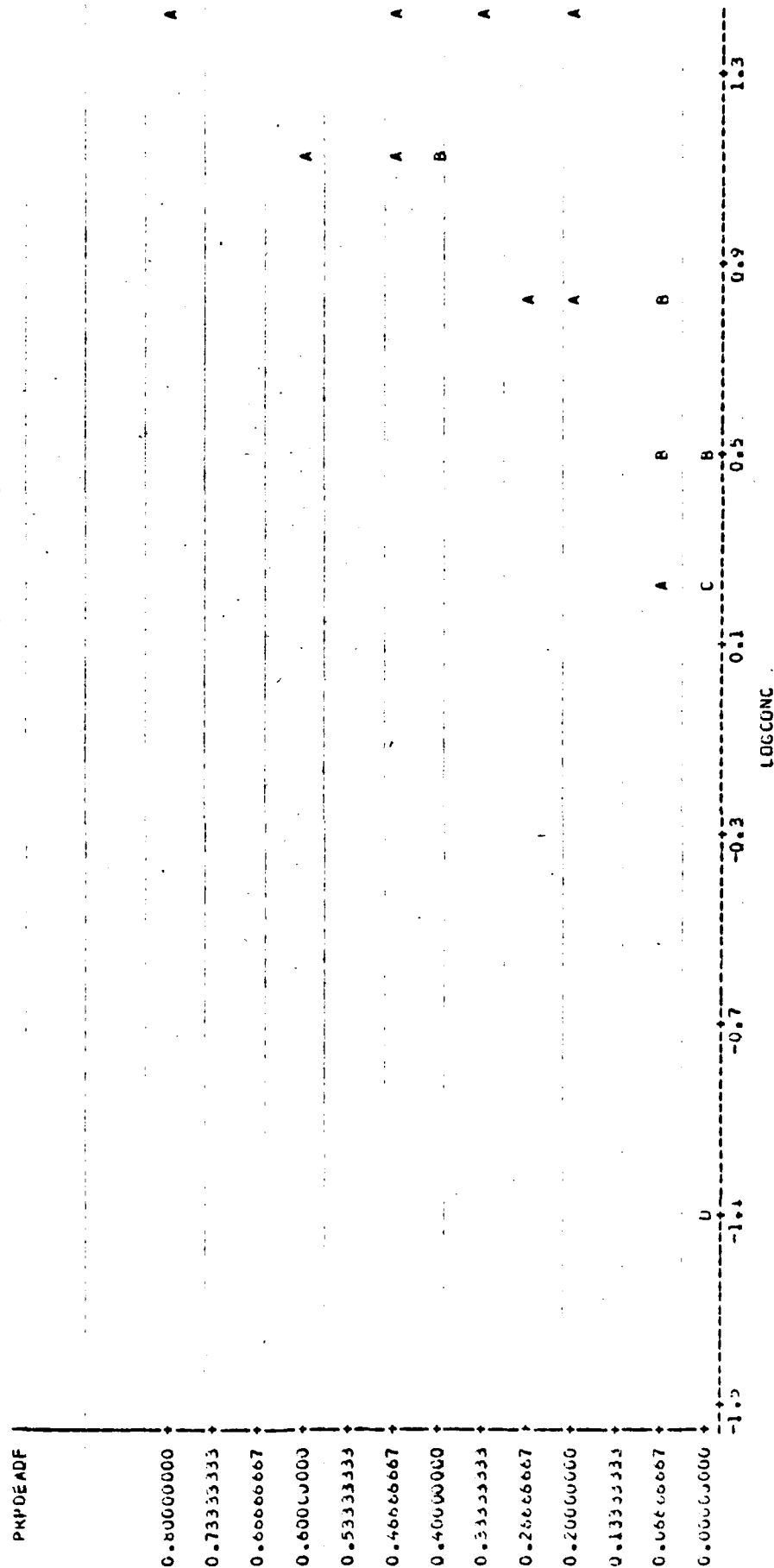


Figure XVI.3

XVII. ANALYSIS OF QUANTITATIVE RESPONSES

Most of the preceeding discussion has been concerned with aspects of the analysis of quantal survival data. This includes preliminary graphical displays, tests for heterogeneity among tanks within groups, outlier detection procedures, adjustments for tank to tank heterogeneity, analysis of variance and multiple comparison procedures, and various types of dose response estimation procedures.

Quantitative responses such as length and weight are also recorded during and at the conclusion of toxicity tests. Weight measurements (in mgs) on all surviving fish at the conclusion of early life stage tests are standard. The statistical analysis of such responses is discussed in this section. This discussion is also directly relevant to the statistical analysis of the length and weight measurements that are taken at periodic intervals (e.g. at 30 day or 60 day intervals) in full life cycle tests.

The approach to analyzing weight and length data is directly analagous to the approach to analyzing survival data. Each aspect of analyzing quantal data, mentioned above, has a direct counterpart for analyzing quantitative data. In fact the procedures appropriate for analyzing quantitative responses are more "standard" and more familiar to most users of statistical methodology than those appropriate for analyzing quantal responses.

In addition to weight and length data a quantitative response often recorded and analyzed in mammalian toxicology tests is time to death (or time to tumor or time to any apriori specified event). Such time to death data provide more information than the 30 day, 60 day, 120 day, etc., survival rates that are commonly reported in aquatic toxicity tests. In particular knowledge of time to death of each embryo or fry yields the percent survival responses as a byproduct. However 30 day survival data will not reveal whether the fish that died did so on day 1, day 15, or day 29. Such information is important for understanding the mechanisms by which the toxicants act.

The analysis of time to death data involves working with censored responses. Parametric and nonparametric approaches to the analysis of such data are discussed in a number of books such as Gross and Clark [49] and Kalbfleisch and Prentice [50].

Time to death (to the nearest day) data is usually collected as part of the day to day test procedure since the tanks are examined daily (or at least on weekdays) for dead fish and these are removed and recorded. Unfortunately, time to death data is not routinely reported as part of the experimental results. In particular time to death of individual fish was not included in any of the data sets made available to us. As such, the analysis of censored life data is not discussed here.

We would make a strong recommendation that time to death of individual fish be routinely reported in the future instead of or in addition to 30 day, 60 day, etc., percent survival data. This would require little additional cost or effort but could possibly provide valuable additional information.

In the remainder of the section we consider the various aspects of analyzing the weights recorded on survivors of 30 day early life stage tests. Before getting down to the technical and methodological issues, several conceptual points should be discussed. The primary difficulty in the interpretation of weight data is the confounding of weight gain with survival. A number of scenarios can be postulated leading to different conclusions about the relationships to be anticipated. Death can be thought of as the first order effect of the toxicant and weight loss (or lesser weight gain) as a secondary effect. If death and weight loss represent different degrees of severity of the same mechanism then one would expect that average weights of survivors would decrease as the mortality rate increases and lack of observation of such a decrease might be interpreted as lack of effect. However, since weights are measured just on survivors a selection phenomenon may be occurring. Presumably, in the various treatment groups the stronger fish survive while the weaker fish die. Presumably the weaker fish would have gained less weight on average than the stronger fish, had they survived (e.g., if they were in the control group). Since greater numbers of weak fish survive in the control and low treatment level groups than in the higher treatment level groups, these weak survivors might decrease the average weight gain relative to the strong survivors in the treatment groups. Thus an increase in observed average weight with treatment level might be possible, or if the toxicant reduces weight gain in the treatment groups, the selection and reduction effects may offset one another, thereby resulting in no observable trend. Therefore the biological meaning of observed trends in weight gain with increasing concentration or the lack of observed trends depends very much on the biological assumptions about toxicant mechanisms and about association between survival and weight loss.

One way to reduce or eliminate the confounding of the survival and weight gain responses is to confine weight gain comparisons to those concentration groups whose mortality rates are not significantly (either biologically or statistically) greater than that in the control groups. The rationale for this viewpoint is that mortality is a first order effect while weight gain is a second order effect. Thus in groups with significant mortality, the question of reduction in weight gain is not of concern. Only when the mortality rate approximates that in the control group is the comparison of weight gains important.

We illustrate the analysis of the weight responses with fry data from the Holcombe and Phipps test on compound D. Recall that the observed mortality rates in the control group and the five treatment groups were 0.06, 0.08, 0.08, 0.13, 0.79, 1.00. It was shown in Section XII that the mortality rates in groups 5 and 6 differ significantly from that in the

control group and group 4 is borderline (statistically) significant according to Williams' procedure. Thus it might be reasonable to confine weight gain comparisons with the control group to treatment groups 2, 3, and possibly 4.

A. Preliminary Scatterplots

A number of plots were prepared showing the means and standard deviations of weight by tank plotted against group number, concentration, log concentration, proportion of survivors, and other variables. Several of these plots are shown in Figures XVII.1 to XVII.4. Figure XVII.1 shows average weight per tank vs group number. Group 6 does not appear in this plot since it had 100 percent mortality. A downward trend in average weight with increasing group number is evident. Note that the total range of variation in average weights is not very great -- between 100 mg and 145 mg. Figure XVII.2 shows standard deviation of weight per tank vs group number. With the exception of one tank in group 5 there appears to be no trend in standard deviation with group. Note that the standard deviation estimates in group 5 are less stable than those in the other groups since they are based on many fewer observations. Figures XVII.3, XVII.4 show the average weight per tank vs concentration and vs log concentration respectively (more specifically $\log(1 + \text{CONC})$). The decreasing trend in average weight with increasing concentration is again evident. In Figure XVII.4 a linear or quadratic trend in log concentration can be seen among the treatment groups.

B. Outlier Detection Procedures and Testing for Tank to Tank Heterogeneity Within Groups

Analysis of variance models were fitted to the individual weights and logarithmic weights to determine if there is any statistical evidence of tank to tank heterogeneity within groups or of differences in average weights across groups. The two way mixed model

$$W_{ijk} = \mu + \alpha_i + \tau_{j(i)} + \epsilon_{ijk} \quad \begin{array}{ll} i = 1, \dots, I & j = 1, \dots, J \\ k = 1, \dots, n_{ij} \end{array}$$

was specified where W_{ijk} corresponds to the weight or to the log weight of the k -th fish within the j -th tank of the i -th group, α_i is the fixed group effect, $\tau_{j(i)}$ is the random effect of the j -th tank within the i -th group, and ϵ_{ijk} is the experimental variation. It is assumed that $\tau_{j(i)}$ are independent $N(0, \sigma_\tau^2)$ and ϵ_{ijk} are independent $N(0, \sigma_\epsilon^2)$ and the $\{\tau_{j(i)}\}$, $\{\epsilon_{ijk}\}$ are independent. In the case of the Holcombe and Phipps Compound D fry mortality data $I = 5$, $J = 4$, n_{ij} varies with tank and with group but is nearly constant in groups 1 to 4. The model was fitted to the data using PROC GLM in the SAS statistical computing system [12]. The results are shown in Figures XVII.5 to XVII.7.

Figures XVII.5 and XVII.6 show the analysis of variance tables for the responses weight and log weight respectively. Figure XVII.7 shows the expected mean squares of the entries in the analysis of variance tables in Figures XVII.5 and XVII.6. The conclusions from the fits to the untransformed weights and to the logarithmic weights are very similar. From the expected mean square of TANK (GRP) indicated in Figure XVII.7 it is seen that the hypothesis $H_0: \sigma_T^2 = 0$ is tested by comparing the TANK (GRP) mean square with the error mean square. The resulting F tests are nonsignificant for both the untransformed and logarithmic weight responses (observed significance levels 0.64, 0.70 respectively). The estimated variance components are:

untransformed weights

$$\hat{\sigma}_T^2 = \left(\frac{11324.61}{15} - 900.35 \right) / 19.19 = -7.58$$

logarithmic weights

$$\hat{\sigma}_T^2 = \left(\frac{1.136}{15} - 0.097 \right) / 19.19 = -0.001$$

Thus in both cases $\hat{\sigma}_T^2$ is set equal to zero. Therefore in this example the error mean square may be used as an error yardstick against which to compare the fixed effect mean squares for group effects.

In general σ_T^2 will not be equal to zero and so an appropriate error yardstick will be a linear combination of the error mean square and the tank (group) mean square. To see how this works consider the test of the hypothesis $H_0: \alpha_1 = 0$, that is no group effects. This null hypothesis is obviously false and the test given by PROC GLM, based on the error mean square with 366 d.f. rejects H_0 very strongly. However the error mean square underestimates the variability of the group mean square if σ_T^2 is greater than zero. The type IV expected mean square for group is shown in Figure XVII.7 to be $\sigma_e^2 + 18.2048\sigma_T^2$. Thus the tank (group) mean square with 15 d.f. is a more appropriate error yardstick than the error mean square estimates σ_e^2 . In general a linear combination of these two mean squares would be an even better yardstick.

The classical approach to combining expected mean squares is based on choosing that linear combination which yields an unbiased estimator of $\sigma_e^2 + 18.2048\sigma_T^2$. Namely

$$w[\sigma_e^2 + 19.1915\sigma_T^2] + (1 - w)\sigma_e^2 = \sigma_e^2 + 18.2048\sigma_T^2$$

Thus

$$w = \frac{18.2048}{19.1915} = 0.95$$

Note that w does not depend on σ_T^2 .

If the design had been completely balanced, this would have led to using the tank (group) mean square with 15d.f. irrespective of the value of σ_T^2 . Such an approach is analagous to carrying out analyses on a per tank basis rather than on a per fish basis.

For the untransformed weights

$$.95 \text{ MS TANK (GRP)} + .05 \text{ MS ERROR} = .95(754.97) + .05(900.35) = 762.26$$

For the logarithmic weights

$$.95 \text{ MS TANK (GRP)} + .05 \text{ MS ERROR} = .95(0.0757) + .05(.0972) = 0.0768$$

To calculate the effective number of degrees of freedom of this linear combination, assume that

$$.95 \text{ MS TANK (GRP)} + .05 \text{ MS ERROR} \sim (\sigma_e^2 + 18.2048\sigma_T^2)\chi_v^2/v$$

where v is unknown. Equating the variances of the two sides we obtain:

$$(.95)^2(\sigma_e^2 + 19.1915\sigma_T^2)^2 \frac{2}{15} + (.05)^2\sigma_e^2 \frac{2}{366} = (\sigma_e^2 + 18.2048\sigma_T^2)^2 \frac{2}{v}$$

Approximating the expectations by the mean squares yields

(a)untransformed weights

$$(.95)^2(754.97)^2 \frac{1}{15} + (.05)^2(900.35)^2 \frac{1}{366} = (762.26)^2 \frac{1}{v}$$

or $v = 16.9$

(b)logarithmic weights

$$(.95)^2(.0757)^2 \frac{1}{15} + (.05)^2(.0972)^2 \frac{1}{366} = (.0768)^2 \frac{1}{v}$$

or $v = 17.1$

Comparison of the type IV group mean square with this error yardstick yields

untransformed weights: $7649.96/762.26 = 10.04$

logarithmic weights: $0.6743/0.0768 = 8.78$

Although in the case of unbalanced data the numerator and denominator mean squares are not strictly independent and the denominator "mean square" is not strictly distributed as chi square, an approximate test is usually constructed by treating these ratios as F ratios with degrees of freedom 4 and 17. These ratios are of course very highly statistically significant according to this yardstick. Thus there is very strong statistical evidence of differences in average weights among groups.

It should be noted that this approach of using essentially the tank (group) mean square with 15d.f. as error yardstick is very close to carrying out group to group comparisons on a per tank basis rather than on a per fish basis. Figure XVII.8 shows the output from a one way analysis of variance to compare group means using the tank means as basic input data. This corresponds to a quantification of the relations seen in Figure XVII.1. Again the group effects are highly significant. The expected mean square for tank (group) as shown in Figure XVII.7 is as if each tank contained 19.19 fish on average. (The average is actually 19.3). If we divide MS TANK (GRP) by 19.19 we obtain 39.34 which agrees quite well with the error mean square of 39.40 in Figure XVII.8.

An alternative approach to pooling mean squares in the analysis of variance is based on finding that linear combination of tank (group) and error mean squares which minimizes the mean square difference from $\sigma^2 + 18.2048\sigma_t^2$. Namely choose w so that

$$E[w \text{ MS TANK (GRP)} + (1 - w) \text{ MS ERROR} - (\sigma_e^2 + 18.2048\sigma_t^2)]^2$$

is minimized. The resulting choice of w will depend on the relative magnitudes of σ_e^2 and σ_t^2 . As σ_t^2/σ_e^2 approaches zero, more and more emphasis will be placed on the error mean square because its reduced variance will more than compensate for its bias. The above expectation can be calculated to be

$$w^2(\sigma_e^2 + 19.19\sigma_t^2)^2 \frac{2}{15} + (1 - w)^2 \sigma_e^4 \frac{2}{366} + [(19.19w - 18.2048)\sigma_t^2]^2$$

We wish to choose $w(0 \leq w \leq 1)$ to minimize this expression. If we substitute in the estimates of the mean squares and of the variance components we obtain

$$w^2(754.97)\frac{2}{15} + (1 - w)^2(900.35)\frac{2}{366} + 0 = 75997.3w^2 + 4429.7(1 - w)^2$$

which has its minimum in the interval $0 \leq w \leq 1$ at $w = 0.055$. Thus the second approach leads to the error yardstick

$$.055 \text{ MS TANK (GRP)} + .945 \text{ MS ERROR} = 892.35$$

with approximate degrees of freedom obtained by solving the equation

$$(.055)^2 (754.97)^2 \frac{1}{15} + (.945)^2 (900.35)^2 \frac{1}{366} = (892.35)^2 \frac{1}{v}$$

or $v = 380$.

Thus this allocation effectively leads to the use of the error mean square in this case and is very different from that obtained by equating expected mean squares. This alternative pooling scheme is most useful when there are an inadequate number of degrees of freedom for estimating MS TANK (GRP) because it then puts more weight on the error mean square. The criterion used does not of course, insure that the resulting mean square is an unbiased estimate of $\sigma_e^2 + 18.2048\sigma_\gamma^2$.

It should be noted that the GLM procedure permits the decomposition of the model sum of squares into individual degree of freedom components. This feature is illustrated in Figure XVII.8. The linear and quadratic components of trend are defined by the contrasts $(-2, -1, 0, 1, 2)$ and $(2, -1, -2, -1, 2)$ respectively. These contrasts single out physically important comparisons among the groups to test and estimate and thereby increase the sensitivity of the analysis of variance tests. This approach is analogous to carrying out a one sided measure of association test with qualitative survival data.

The residuals from the analysis of variance fits can be used to check distributional assumptions and to detect outliers. Figures XVII.9 - XVII.12 display the arithmetic and logarithmic residuals from the fits in Figures XVII.5 and XVII.6 respectively. Figures XVII.9 and XVII.10 show the residuals plotted vs group. No outliers are obvious. The variability seems constant with group. The residuals from the untransformed weights appear to be much more symmetric about zero than those from the logarithmic weights. Figures XVII.11 and XVII.12 show normal probability plots of the arithmetic and logarithmic residuals respectively. The plot in Figure XVII.11 looks much more nearly normal than that in Figure XVII.12.

The lowest two residuals in Figure XVII.11 lie below the line through the remainder of the data. To determine whether there is any statistical evidence that these observations are outliers we can test whether the most extreme of 386 independent normally distributed random variables with mean 0 and standard deviation 30 is likely to exceed 118 in absolute value. (The two extreme residuals correspond to observations 209 and 220, are from group 3, tanks A and B, have values -112.9 and -117.9, and are associated with fish having reported weights of 15 mgs and 10 mgs. I am assuming that these are the correct weights, but this should be checked).

$P[\text{most extreme of 386 observations greater than 118 in absolute value}] =$

$$1 - [P(-118 < X < 118)]^{386} = 1 - [2\Phi\left(\frac{118}{30}\right) - 1]^{386} = 1 - (.999950)^{386} \\ = 1 - .98 = .02$$

Thus there is statistical evidence that this extreme residual does not conform to the others. Whether this represents a clerical error or natural biological variation would need to be determined.

Assuming that the extreme observation is an outlier, the second most extreme observation can be compared to the extreme of 385 observations.

$P[\text{most extreme of 385 obsvns greater than 113 in absolute value}] =$

$$1 - [P(-113 < X < 113)]^{385} = 1 - [2\Phi\left(\frac{113}{30}\right) - 1]^{385} = 1 - (.99990)^{385} \\ = 0.037$$

There is thus statistical evidence that this second most extreme observation is also an outlier.

Basic records should be examined to determine if these observations are valid. If not, they should be corrected or deleted and the modified data reanalyzed. If they in fact represent natural biological variation then biological judgement should be used to determine whether or not to retain these observations with the remainder.

C. Multiple Comparison Procedures and Regression Analyses

Based on the results of the analysis of variance calculations previously discussed, we can carry out comparisons of average weight gains across groups. The average weight gains and numbers of animals per group are:

Group	1	2	3	4	5
N	94	92	92	87	21
Average	131.2	135.59	127.9	113.1	108.1

The standard errors of these averages based on averaging the responses from four tanks and varying numbers of fish per group, are $1/4(\sigma_T^2 + \sigma_e^2/23.5)$, $1/4(\sigma_T^2 + \sigma_e^2/23)$, $1/4(\sigma_T^2 + \sigma_e^2/23)$, $1/4(\sigma_T^2 + \sigma_e^2/21.75)$, $1/4(\sigma_T^2 + \sigma_e^2/5.25)$. The variance components σ_T^2 , σ_e^2 can be estimated by appropriate linear combinations of the tank (group) and error mean squares displayed in Figure XVII.5. For the Holcombe and Phipps fry mortality data $\hat{\sigma}_T^2 = 0$ and so we can estimate the standard errors using the error mean square.

Thus

$$\hat{\sigma}_e^2 = 900.355 \quad \text{with 366 d.f.}$$

Alternative standard error estimates with appropriate degrees of freedom can be constructed using approaches analogous to those discussed in the previous subsection.

We apply Williams' procedure [36] to determine which groups have (statistically) significantly lower weight gain than the control group. We first need to adjust these mean values so that they are in monotone decreasing sequence. We simply calculate weighted average $(94 \times 131.2 + 92 \times 135.59)/186 = 133.37$ of the averages in groups 1 and 2. The modified averages are now in monotone decreasing order. We declare the group i average weight gain to be significantly smaller than the control average if

$$\bar{X}_{i,adj} - \bar{X}_1 < -\bar{t}_e (1/N_i + 1/N_1)^{1/2}$$

Note that $\bar{X}_{i,adj}$ is the adjusted average whereas \bar{X}_1 is the unadjusted average. We use the factor \bar{t} obtained from Williams' table, which is derived under the assumption of equal group sample sizes. This assumption is quite reasonable for groups 1-4. We assume that $\hat{\sigma}_e^2$ is estimated with 366 d.f. A more conservative assumption might be use 15 d.f. since this is the amount of information concerning σ_e^2 . This would raise \bar{t} from 1.75 to 1.88.

The group i mean is declared to be significantly lower than the control mean if

$$\bar{X}_{i,adj} < \bar{X}_1 - \bar{t}_e (1/N_i + 1/N_1)^{1/2} = 131.2 - (1.75)(30.01)(1/N_i + 1/N_1)^{1/2}$$

$i = 2:$	$\bar{X}_{2,adj} = 133.37$	critical value = 123.50
$i = 3:$	$\bar{X}_{3,adj} = 127.9$	critical value = 123.50
$i = 4:$	$\bar{X}_{4,adj} = 113.1$	critical value = 123.39
$i = 5:$	$\bar{X}_{5,adj} = 108.1$	critical value = 118.52

Thus the average weight gains in groups 4 and 5 are declared by Williams' procedure to be (statistically) significantly lower than the control group average at the 5 percent level. With respect to mortality, group 5 is obviously much different than the control group and group 4 is borderline (statistically) significantly different. Thus the quantal survival and quantitative weight responses yield essentially the same conclusions.

We can fit regression models to the weight gain data to quantify the trends in averages across groups. Figure XVII.4 shows average weight gain plotted against $\log(1 + \text{CONC})$. The responses in the four treatment groups show a definite trend, mostly linear but possibly with some second order curvature. Figure XVII.13 shows the results of fitting a cubic

polynomial in $\log(1 + \text{CONC})$ to the treatment groups and an indicator function to the control group. Namely the model

$$W_{ijk} = \beta_0 + \beta_1 \text{ICTL} + \beta_2 \text{LCONC} + \beta_3 \text{LCONC}^2 + \beta_4 \text{LCONC}^3 + \epsilon_{ijk}$$

was fitted to the weight data where $\text{ICTL} = 1$ if $\text{GRP} = 1$ and 0 otherwise and $\text{LCONC} = \log(1 + \text{CONC})$. This model fits a cubic polynomial to the treatment groups. The parameter β_1 represents the difference between the control group mean and the extrapolation back to $\text{LCONC} = 0$ along the cubic polynomial. The contrasts estimated at the bottom of the figure correspond to the differences between the mean responses at the treatment groups, based on the polynomial fit, and the control group response.

A complication in inference procedures arises if there is tank to tank heterogeneity within groups. Observations within the same tank are then dependent due to a common tank effect. The variation of the type IV mean square for GRPS is seen in Figure XVII.7 to be inflated from σ_e^2 to $\sigma_e^2 + 18.2048\sigma_t^2$ due to such heterogeneity. The standard errors in Figure XVII.13 might then be inflated by the factor $[1 + 18.2048\sigma_t^2/\sigma_e^2]^{1/2}$ to account for such heterogeneity. The quantity $\sigma_e^2 + 18.2048\sigma_t^2$ can be estimated by pooling mean squares in Figure XVII.5 in at least one of two different ways, as discussed in the previous subsection. These yields estimates of 726.26 with 17 d.f. or 892.35 with 380 d.f. Alternatively, regression fits can be carried out on a per tank basis, as is commonly done. This turns out to be very similar to using the 17d.f. variance estimate.

Since σ_t^2 was nonsignificant in the previous ANOVA fit, since $\hat{\sigma}_t^2 = 0$, and since the variance estimate with 380 d.f. is very similar to the error mean square in Figure XVII.13, we will use the error mean square as the basis of standard error calculations in this example. It should be noted however that this is appropriate only if $\sigma_t^2 = 0$.

We see from the type I sums of squares in Figure XVII.13 that the linear component of trend is highly significant while the quadratic and cubic trends are nonsignificant over and above the linear trend. This agrees with the appearance of Figure XVII.4. The quadratic and cubic terms should be deleted and the model refitted. The contrasts at the bottom of Figure XVII.13 show, in agreement with the results from Williams' procedure, that groups 4 and 5 differ significantly from the control group while groups 2 and 3 do not.

QUALITY OF WEED DATA (MUST) FROM SURVIVORS OF HOLCOMBE AND PHELPS TEST ON COMPOUND B
 (CATTLE) REPLY OF ADAMS, THE LEVALLY GROUP, MORTALITY, CONCENTRATION - 20:19 TUESDAY, APRIL 7, 1981

PLOT TO WINDWARD LEGEND: A = 1 ORS, B = 2 ORS, ETC.

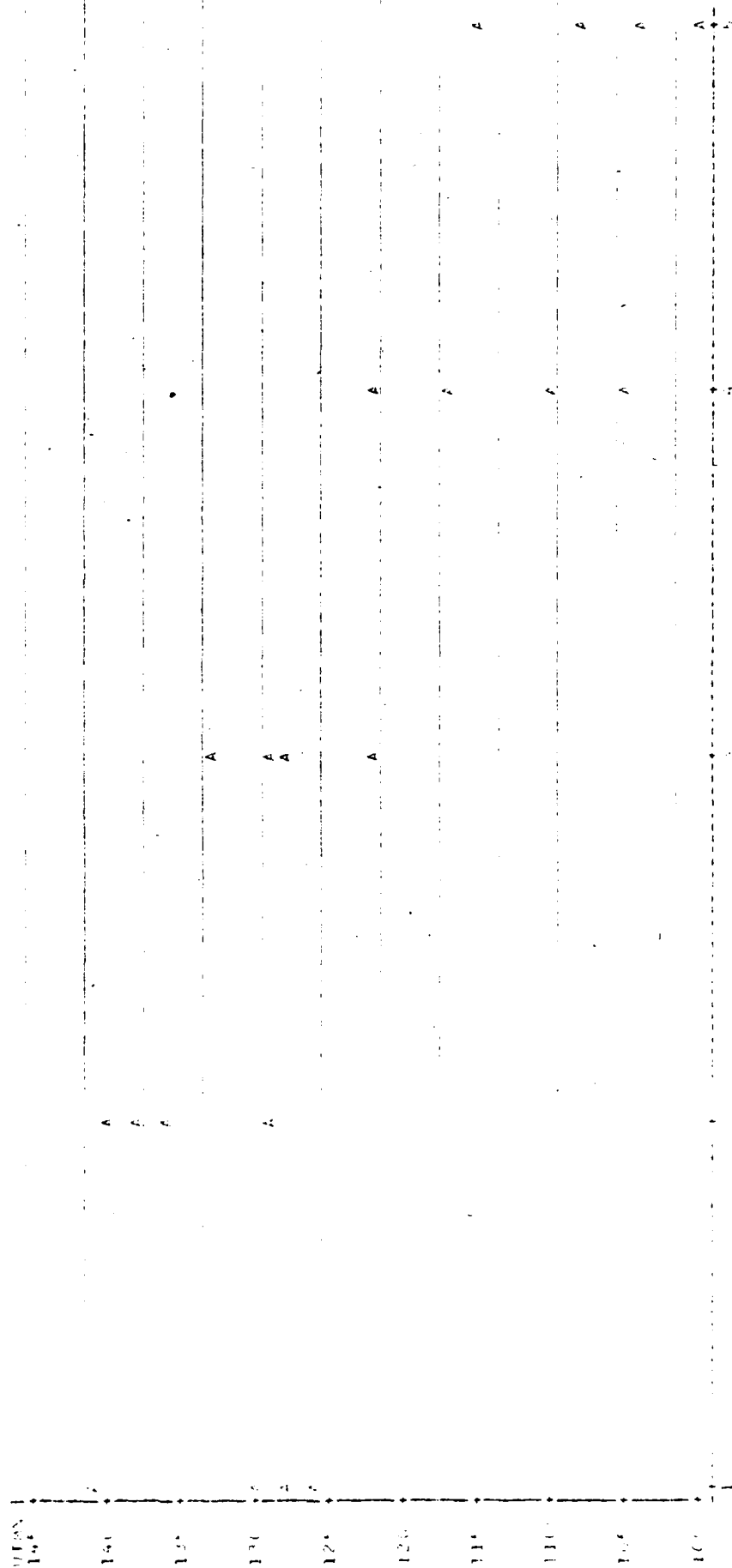


Figure XVII.1 Average weight per tank vs group

ANALYSIS OF WEIGHT DATA(S) FROM SURVIVORS OF HOLCOMBE AND PHILIPS
 ECOTRIPPLETS OF MEANS, STD DEVS BY GROUP, MORTALITY, CONCENTRATION
 PLANT OF MISTRESS. LEGEND: A = 100%, B = 200%, C = 300%, ETC.

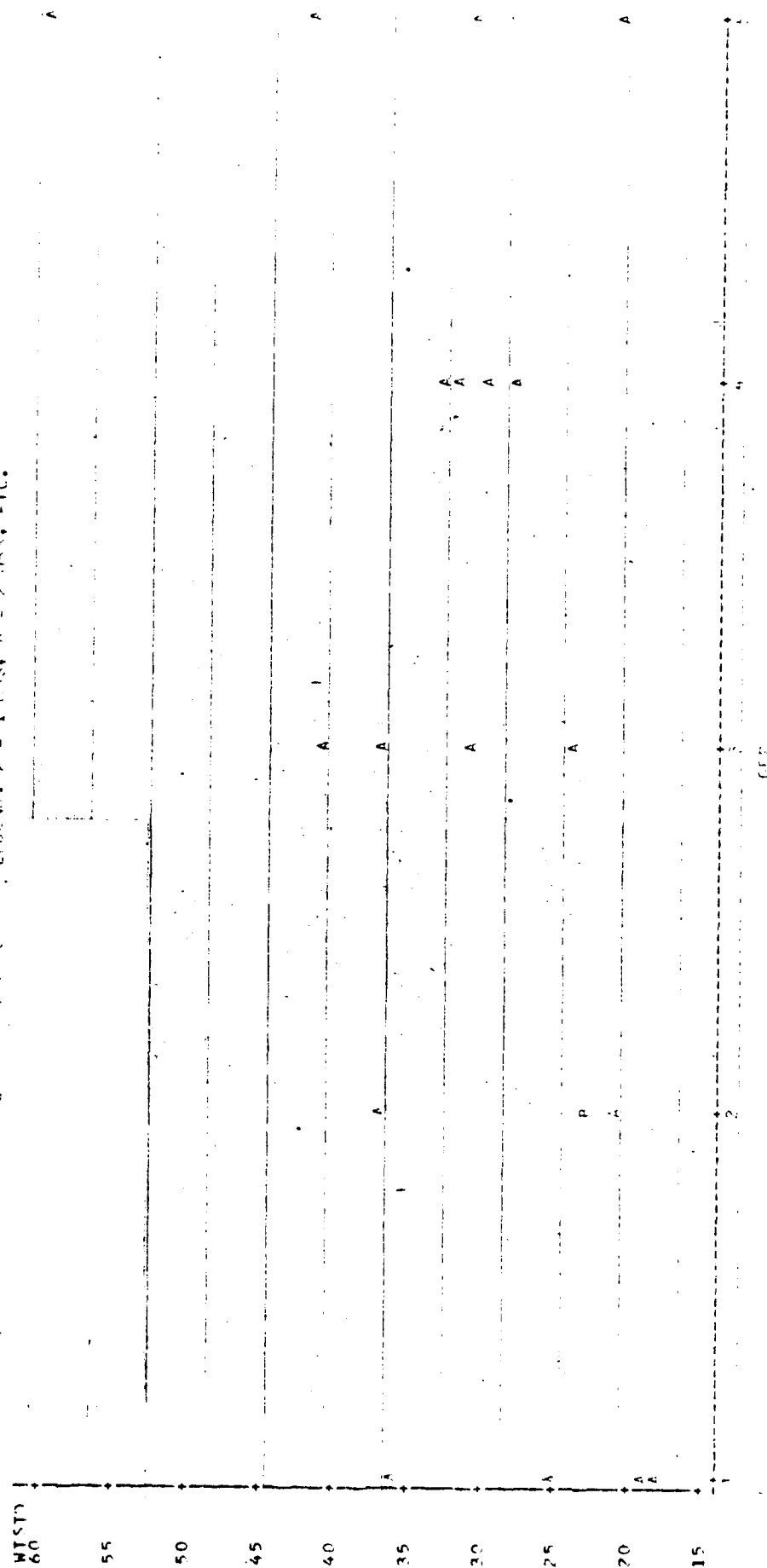


Figure XVII.2 Standard deviation of weight per tank vs group

ANALYSIS OF SUBJECT DATA(S) FROM SURVIVORS OF POLIO AND POLIOUS TEST ON COMPOUND 6
 QUANTITIES OF MEANS, STD DEVS BY GROUP, MORTALITY, CONCENTRATION 2019 JUL SUAY, APRIL 7, 1981
 PLOT OF WINDCONC LEGEND: A = 1 ORS, B = 2 ORS, ETC.

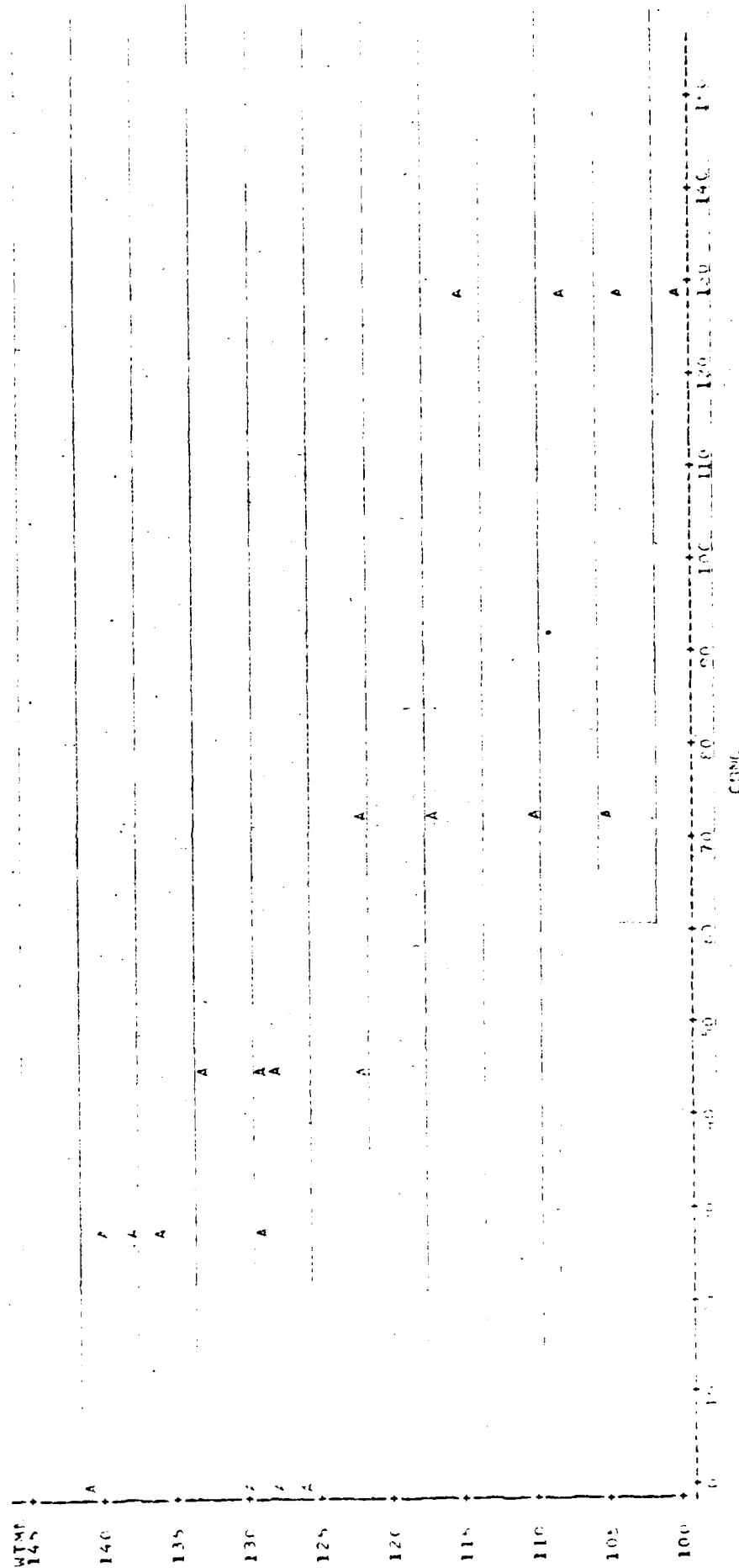


Figure XVII.3 Average weight per tank vs concentration

ANALYSIS OF MICHIGAN DATA (FROM SURVIVORS OF HOLCHESSE AND PHIPPS TEST ON COMPOUND D)
 SCATTERPLOTS OF MEANS, STD DEVS BY GROUP, MORTALITY, CONCENTRATION
 20:19 TUESDAY, APRIL 7, 1961
 PLOT OF WTMPLGNC LEGEND: A = 1 RES, B = 2 RES, ETC.

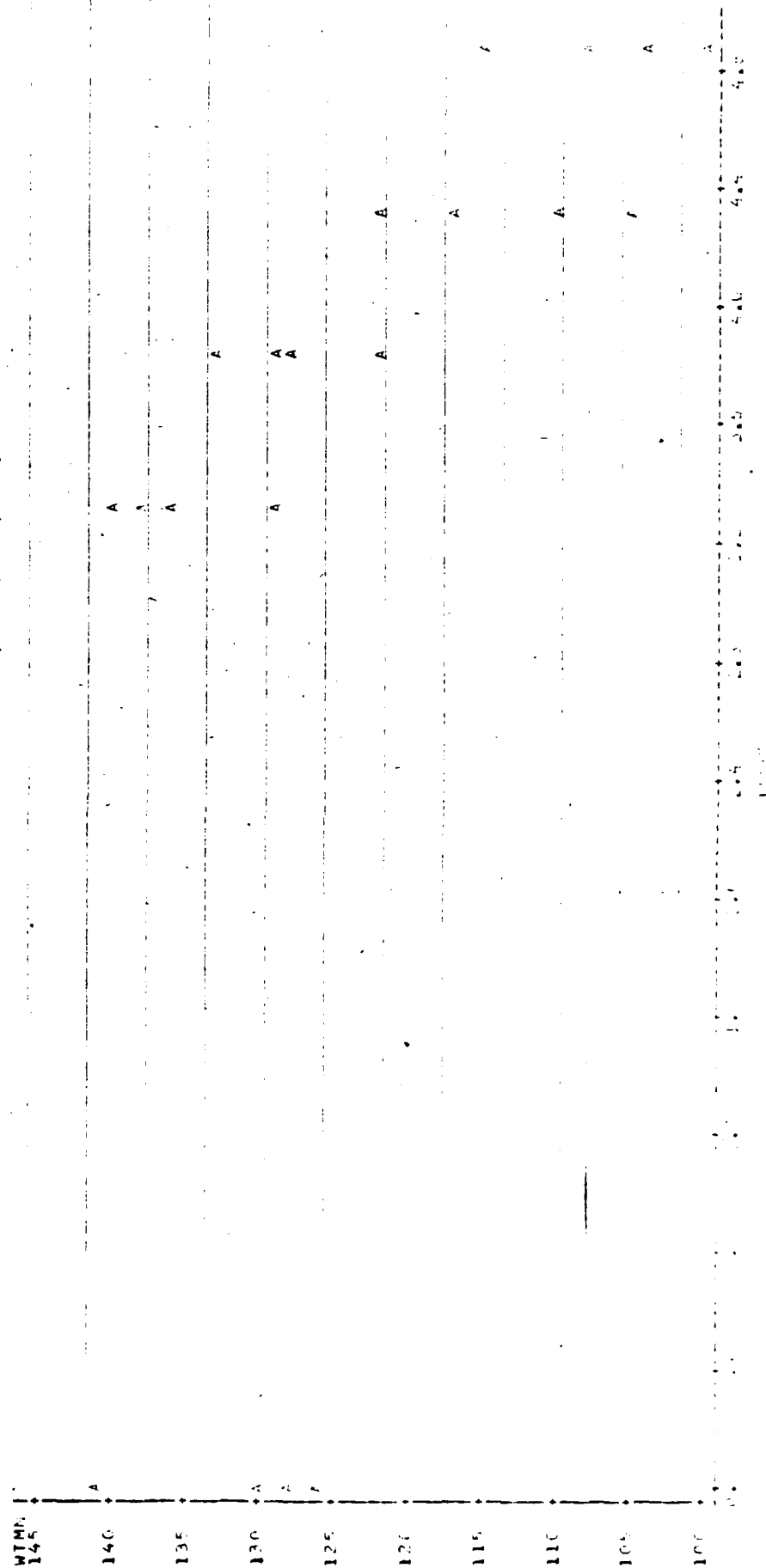


Figure XVII.4 Average weight per tank vs log concentration

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
ANOVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS

TEST ON COMPOUND D
13:38 MONDAY, APRIL 13, 1981

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: WGT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	19	43953.20526379	2313.32659283	2.57	0.0004	0.117682	23.7888
ERROR	366	329529.78955487	900.35461627		STU DEV		WGT MEAN
CORRECTED TOTAL	385	373482.99481865			30.00590969		126.13471503

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
GRP	4	32628.60443368	9.06	0.0001	4	30599.85439173	8.50	0.0001
TANK(GRP)	15	11324.60083011	0.84	0.6350	15	11324.60083011	0.84	0.6250

Figure XVII.5 Analysis of variance fit of individual weights on group and tank within group. Holcombe and Phipps Compound D fry mortality data

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
ANOVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS

TEST ON COMPOUND D

22
13:28 MONDAY, APRIL 13, 1981

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LWGHT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	19	4.02700927	0.21194786	2.18	0.0032	0.101717	6.4986
ERROR	366	35.56349376	0.09716802		STD DEV		LWGT MEAN
CONNECTED TOTAL	385	39.59050303			0.31171782		4.7966309

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
GRP	4	2.89106672	7.44	0.0001	4	2.69707731	6.94	0.0001
TANK(GRP)	15	1.13594255	0.78	0.7015	15	1.13594255	0.78	0.7615

Figure XVII.6 Analysis of variance fit of individual logarithmic weights on group and tank within group. Holcombe and Phipps Compound D fry mortality data

ANALYSIS OF HEIGHT DATAINGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS TEST ON COMPOUND D 20
ANDVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS 13:38 MONDAY, APRIL 13, 1981

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLES: WGT LGHT

SOURCE	TYPE I EXPECTED MEAN SQUARE
GRP	VAR(ERROR) + 19.0254986 VAR(TANK(GRP)) + Q(GRP)
TANK(GRP)	VAR(ERROR) + 19.1914733 VAR(TANK(GRP))
SOURCE	TYPE IV EXPECTED MEAN SQUARE
GRP	VAR(ERROR) + 18.2048497 VAR(TANK(GRP)) + Q(GRP)
TANK(GRP)	VAR(ERROR) + 19.1914733 VAR(TANK(GRP))

Figure XVII.7 Expected mean squares in analysis of variance fits to Holcombe and Phipps fry mortality data

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: WTMN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	2465.21503461	616.30375865	15.64	0.0001	0.806635	5.1030
ERROR	15	590.95730026	39.39715335				WMT. MEAN
CORRECTED TOTAL	19	3056.17233487					123.6012160
						6.27671517	
SOURCE	DF	TYPE I SS	F VALUE	PR > F	TYPE IV SS	F VALUE	PR > F
GRP	4	2465.21503461	15.64	0.0001	2465.21503461	15.64	0.0001
CONTRAST	DF	SS	F VALUE	PR > F			
GRP LINEAR	1	2060.12680161	52.29	0.0001			
GRP QUADRATIC	1	255.60134117	6.49	0.0223			

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > T	STD ERROR OF ESTIMATE
INTERCEPT	106.41666667 B	33.91	0.0001	3.13835759
GRP	24.91983696 B	5.61	0.0001	4.43830786
	29.09329710 B	6.56	0.0001	4.43830786
	21.74452899 B	4.90	0.0002	4.43830786
	7.16726190 B	1.61	0.1272	4.43830786
	0.00000000 B			

NOTE: THE X'X MATRIX WAS GIVEN IN THE MED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

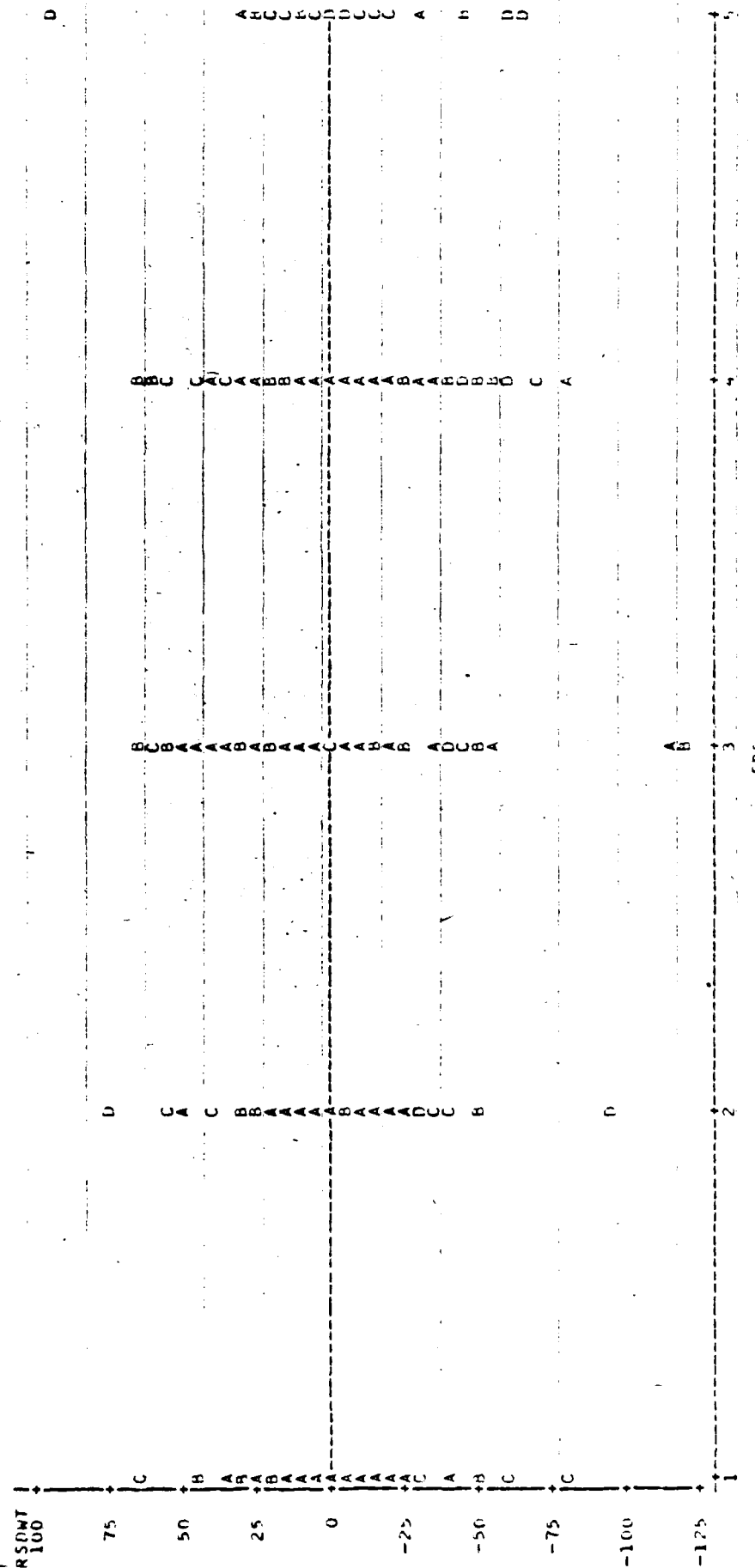
PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > T	STD ERROR OF ESTIMATE
GRP LINEAR	-11.76570911	-7.23	0.0001	9.92435808
GRP QUADRATIC	-29.909994306	-4.55	0.0223	11.74265884

Figure XVII.8 Analysis of variance fit of average weights per tank on group. Holcombe and Phipps Compound D fry mortality data

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
ANOVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS

TEST ON COMPOUND D
13:36 MONDAY, APRIL 13, 1981

PLOT OF RSQNT*GRP SYMBOL IS VALUE OF TANK



NOTE: 276 DFS HIDDEN

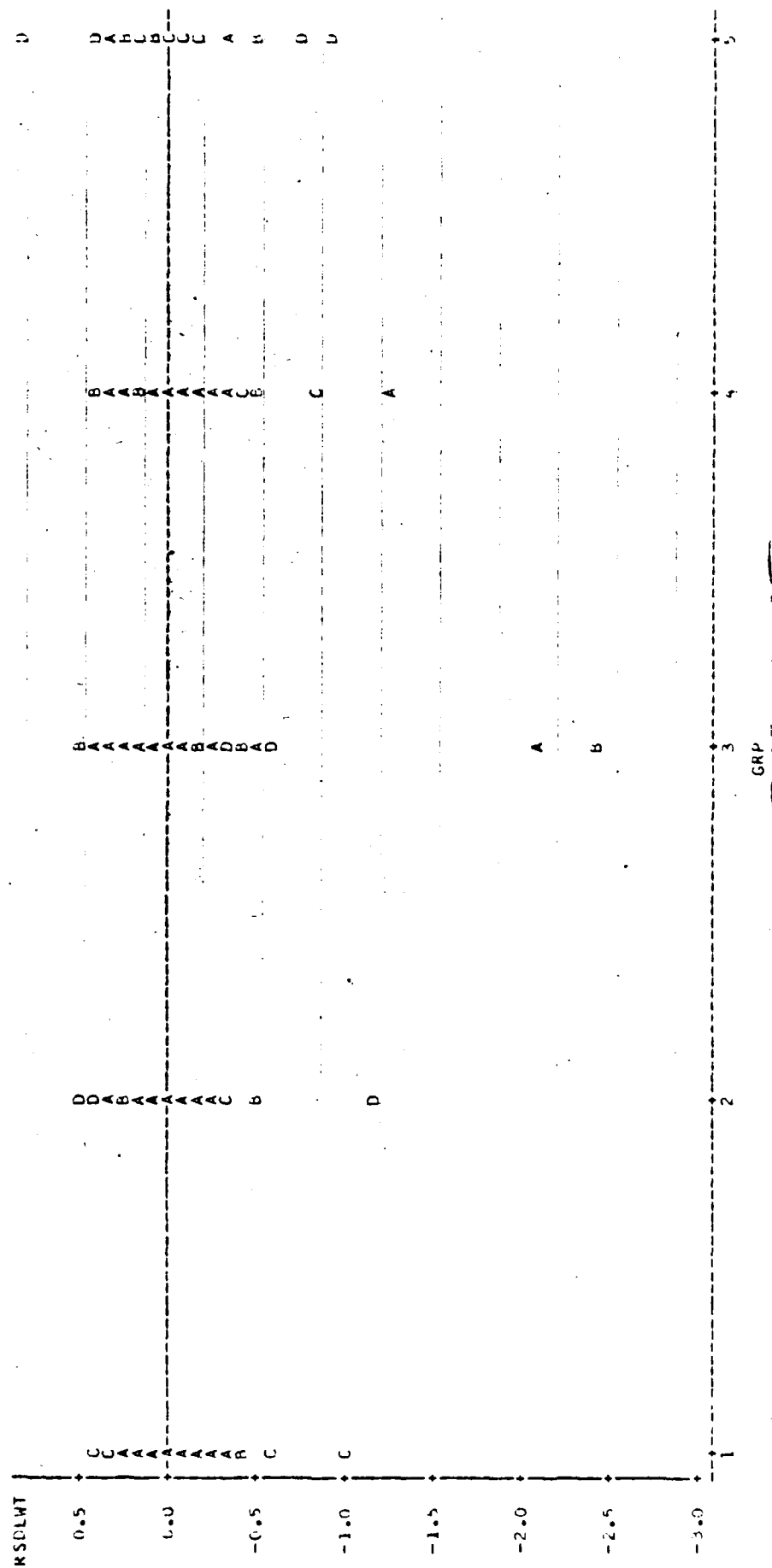
Figure XVII.9 Residuals vs group

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
ANOVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS

TEST ON COMPOUND D
13:38 MONDAY, APRIL 13, 1981

24

PLOT OF RSDLWT*GRP SYMBOL IS VALUE OF TANK



NOTE: 317 LOGS HIDDEN

Figure XVII.10 Logarithmic residuals vs group

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS. TEST ON COMPOUND D
ANDVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS. 13:58 MONDAY, APRIL 13, 1981

PLOT OF RSDWT+RNKRSO LEGEND: A = 1 OBS, B = 2 OBS, ETC.

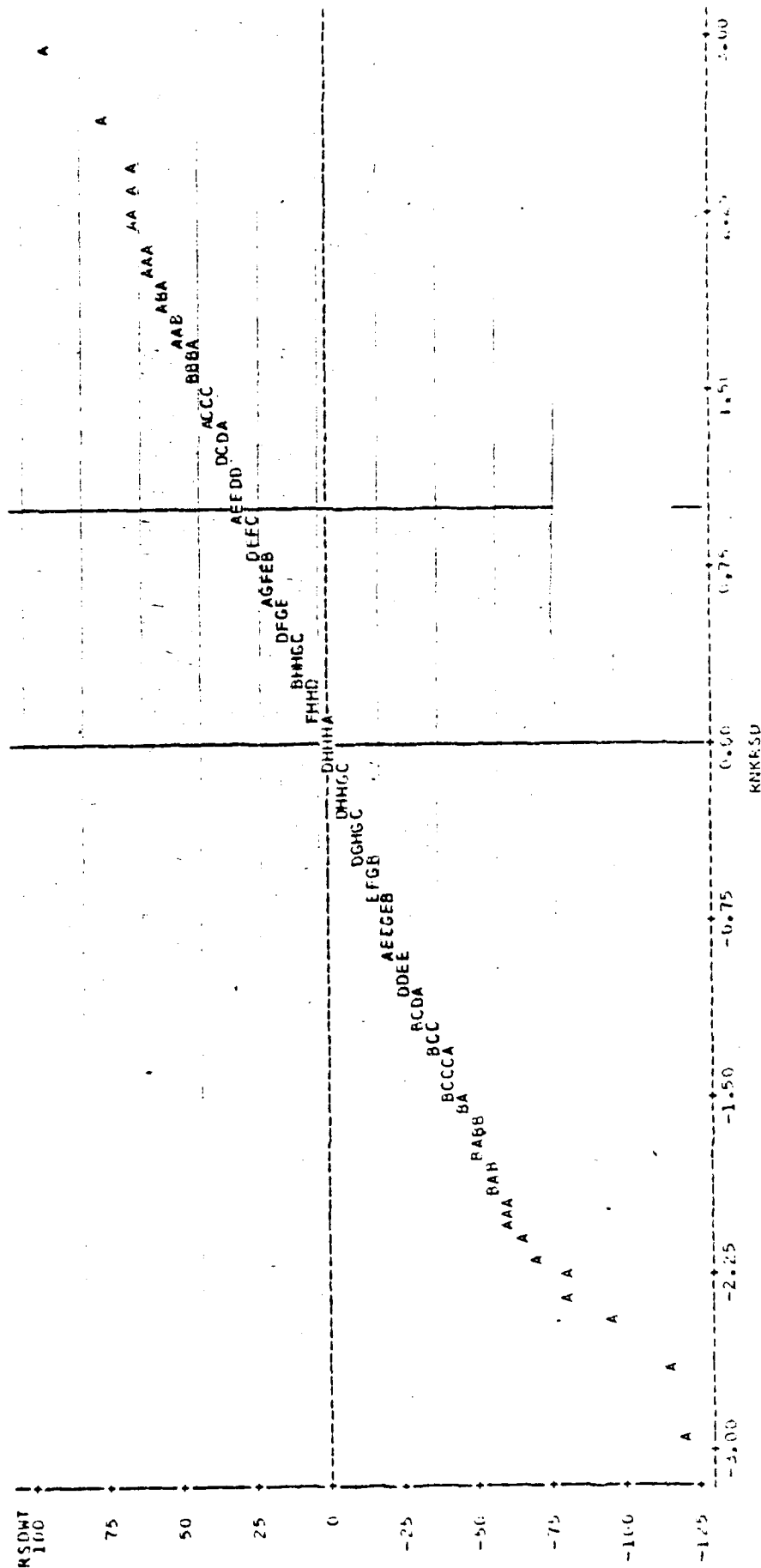


Figure XVII.11 Normal probability plot of residuals from untransformed weights

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
 ANOVA ON ORIGINAL OBSERVATIONS TO DETECT OUTLIERS
 TEST ON COMPOUND D
 13:36 MONDAY, APRIL 13, 1981

PLOT OF RSDLWT*RNKRSDL LEGEND: A = 1 OBS, B = 2 OBS, ETC.

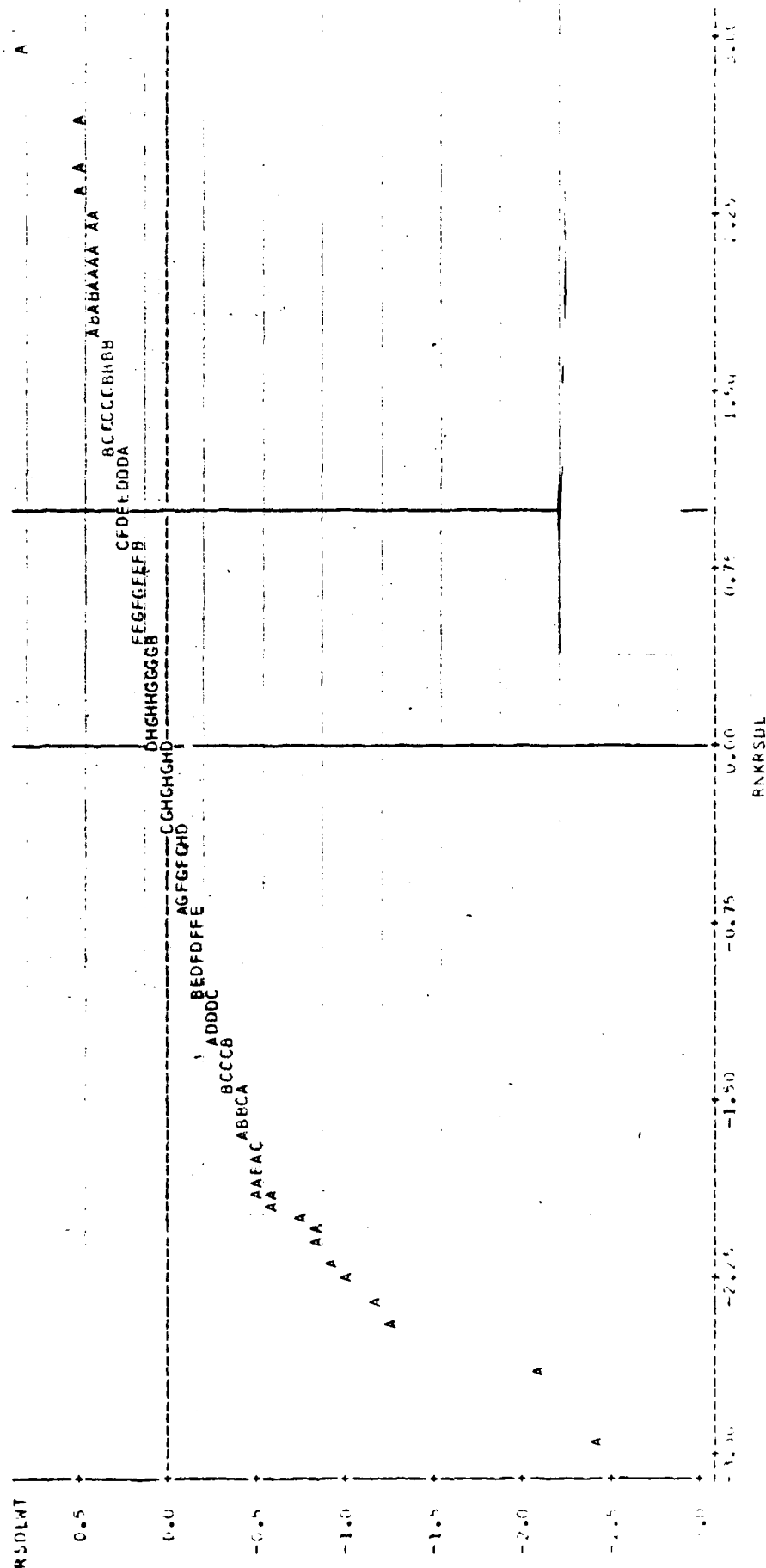


Figure XVII.12 Normal probability plot of residuals from logarithmic weights

AD-A134 309

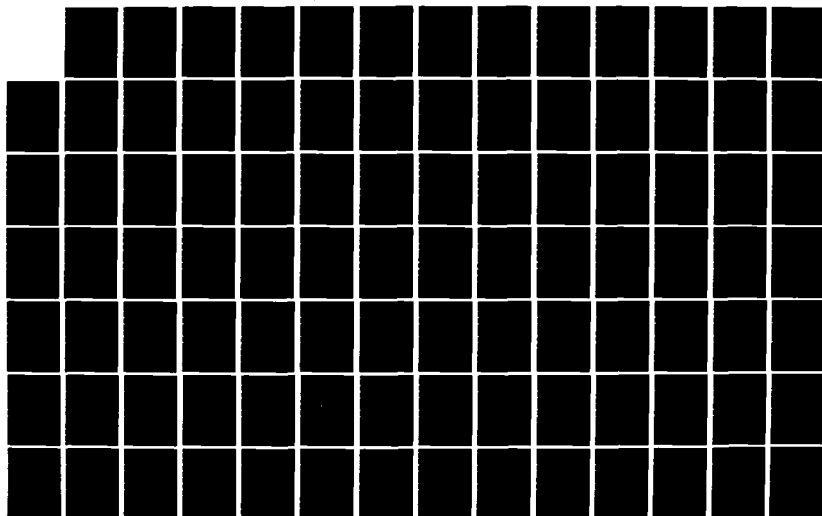
DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF
TOXICITY(U) OHIO STATE UNIV RESEARCH FOUNDATION
COLUMBUS P I FEDER ET AL. MAY 81 DAMD17-79-C-9150

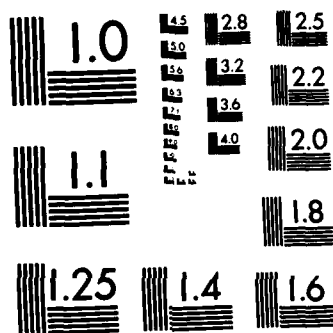
4/5

UNCLASSIFIED

F/G 6/20

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ANALYSIS OF WEIGHT DATA(MGS) FROM SURVIVORS OF HOLCOMBE AND PHIPPS
 SUMMARY STATISTICS BY GROUP

TEST ON COMPOUND D
 21:51 TUESDAY, APRIL 7, 1981

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: WGT

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	22628.60440373	8157.15110093	9.12	0.0001	0.087363	23.7130
ERROR	381	360854.39041492	894.63044597				
CORRECTED TOTAL	385	373482.99481865					

SOURCE	DF	TYPE I SS	F VALUE	PR > F	OF	TYPE IV SS	F VALUE	PR > F
ICTL	1	2244.65330015	3.63	0.0576	1	1171.67790459	1.31	0.2532
LCONE	1	27975.9805225	31.27	0.0001	1	1268.32518700	1.42	0.2345
LCONE*LCONE	1	54.08723345	0.06	0.8059	1	135.11884034	1.54	0.2226
LCONE*LCONE*LCONE	1	1353.88361789	1.51	0.2194	1	1353.88361789	1.51	0.2194

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > T	STD ERROR OF ESTIMATE
INTERCEPT	-1368.88557174	-1.04	0.2970	1310.82853176
ICTL	1500.13172259	1.14	0.2532	1310.83216203
LCONE	1171.76517573	-1.19	0.2345	984.11813656
LCONE*LCONE	-298.13056979	-1.23	0.2226	244.04430528
LCONE*LCONE*LCONE	24.65040278	0.23	0.2194	19.99739889
GRP 2 MINUS CONTROL	4.23042543	0.96	0.2370	4.40146410
GROUP 3 MINUS CONTROL	-3.55365575	-0.77	0.4401	4.47747885
GROUP 4 MINUS CONTROL	-18.22342771	-4.07	0.0001	4.47430192
GROUP 5 MINUS CONTROL	-23.05648719	-3.18	0.0016	7.24410715

Figure XVII.13 Regression fit of weight gain on $\log(1 + \text{CONC})$. Holcombe and Phipps Compound D fry mortality data

XVIII. EXPERIMENTAL DESIGN CONSIDERATIONS

In this section we consider a number of issues pertaining to the design and conduct of aquatic toxicity tests such as precision to be expected as a function of sample size, allocation of tanks among treatment groups, additional variables to measure, numbers of tanks to be run.

A. Assumptions, Additional Variables to Measure, Numbers and Allocations of Tanks

We briefly discuss several assumptions and recommendations associated with planning toxicity tests.

It is assumed that the constraint on size of test is the number of fish tanks that can be run within cost, manpower, and apparatus limitations. The cost of running a test is assumed to be directly proportional to the number of tanks run. Once a tank is put on test, the fish are essentially free. Thus a sample size strategy for early life stage tests is to run as many tanks as can be afforded and fill each tank with the maximum number of embryos or fry that is biologically sensible. This may differ for full life cycle tests.

A sufficient number of fish tanks should be included in the tests to be able to detect the presence of tank to tank heterogeneity within groups. If we have enough degrees of freedom among tanks within groups, the need for pooling mean squares in the analysis of variance in order to improve the sensitivity of tests or estimates of treatment effects is diminished. We can analyze the data on a per tank basis if we wish to, which is appropriate whether or not tank to tank heterogeneity is present. The main difficulty with per tank analyses is when the lack of adequate degrees of freedom diminishes precision of inference. We recommend that there be at least 12 d.f. to estimate tank to tank variation within groups. This would correspond to an average of 3 tanks per group if 6 treatment groups were being run. However to account for the possibility zero percent mortality in the control group and 100 percent mortality in the highest treatment group, it is suggested that 4 replicate tanks per group be run. This would provide 12d.f. for estimating variability just based on the results of the four intermediate groups. A glance at the charts of the noncentral t or noncentral F distribution shows that the power of analysis of variance type tests based on 12d.f. is nearly as great as that with the infinite number of degrees of freedom. For example for $\alpha = 0.05$ and one d.f. in the numerator

noncentrality (θ)	1.5	2.0	2.25	2.5	2.75	3.0
power, 12 d.f.	0.50	0.74	0.83	0.90	0.945	0.972
power, ∞ d.f.	0.56	0.81	0.89	0.94	0.973	0.989

The differences in power are of little qualitative importance.

An important distinction should be recognized between the numbers of tanks needed to estimate the variability of treatment group responses as opposed to the numbers of tanks needed to reduce the tank to tank variability. Suppose that we run J tanks per treatment group, n fish per tank, and that σ_t^2 , σ_e^2 represent the components of variation between tanks and between fish within tanks respectively. Then the variance of a treatment group average is

$$\frac{\sigma_e^2}{nJ} + \frac{\sigma_t^2}{J}$$

If σ_t^2 is large relative to σ_e^2/n , the only way to reduce the variance is by increasing J. However for fixed J, the ability to estimate this variance with 12 d.f. yields nearly as much sensitivity of tests and confidence intervals as an estimate with infinite d.f. This is also reflected in the fact that the upper 97.5 percentile of the t distribution is 2.78 with 4 d.f., 2.18 with 12 d.f., and 1.96 with infinite d.f. Thus 12 d.f., is most of the way between 4 d.f., and infinite d.f.

It should be noted that while we recommend at least 12 d.f., for estimating variability, the tanks do not necessarily need to be equally replicated across groups. In fact we will suggest an unequal allocation later in this section.

Another important aspect of planning experiments is specifying classes of variables. At least six classes of variables can be distinguished: responses to be measured, controlled experimental variables, blocking variables, variables to be held constant, covariates to be measured, and variables to be randomized over. Rather than present a detailed discussion of each class of variable, we will emphasize those aspects which either vary from practice or are less obvious. In addition to the responses currently measured, it was argued in the previous section that individual times to death should be reported. This response must be measured but usually is neither reported nor analyzed. Obvious blocking variables are fish tank or test series. Other, less obvious blocking variables that might be incorporated into investigations are homogeneous subsets of fish (e.g., offspring of common parents, fish raised in the same breeding chamber, fish purchased from a single supplier at a single time, etc), investigators, laboratories, time period when test was conducted, technician, and many others. Some of the latter blocking factors would most naturally occur in round robin tests. In any test program a number of variables are held constant, at least nominally. Examples are water temperature, pH, hardness, levels of additives or impurities; type, amount, and frequency of food; type of fish tank; photoperiod. All these variables must be reported with the experimental results so that experimental conditions can be repeated and results compared across laboratories. Differences in variables held constant will sometimes account for discrepancies

in results. Covariates are factors which cannot be controlled but which can be measured and taken into account when analyzing the data. Covariates commonly reported are deviations from nominal in either controlled experimental variables or in variables held constant. The most obvious covariate is actual test concentration. This should be determined periodically in each tank and reported. The analysis of the data should be based on actual toxicant concentrations rather than nominal. The question of how to summarize toxicant levels has biological as well as statistical aspects. For example is effective level the average, the median, the maximum or some other? The question of frequency of measurement pertains to the short term effects of fluctuations in levels. The greater the effect of short term fluctuations in toxicant levels, the more frequently they need be measured. This aspect will not be considered further here. Other covariates to be measured and used for analysis might be water temperature, hardness, or pH, and measures of the size or health of the brood stock from which the test fish were taken. The remaining class of variables -- those to be randomized over -- is perhaps the most numerous. However the variables thought to be most important were included in the other five categories. These variables, many of which are not explicitly known, are randomized over. Their effects thus enter into the experimental variability. It is hoped that their effects are not too great.

B. Sample Size and Power Considerations for Quantal Survival Data

We first assume that there is no tank to tank variation within groups. We later modify the results to account for tank effects by adjusting sample sizes downward to "effective sample sizes."

If there is no tank to tank variation then it suffices to consider just the number of fish run per control or treatment group. Suppose that there is a control group (group 0) and I treatment groups (groups 1, 2, ..., I). In standard practice $I = 5$. Suppose that we run N_0 fish in the control group and N fish in each treatment group. Then test then involves a total of $N_0 + IN \equiv C$ fish. If we carry out pairwise comparisons of treatment and control groups based on the arcsine transformation of observed response rates, the variances of $2 \arcsin \sqrt{\hat{p}_1} - 2 \arcsin \sqrt{\hat{p}_0}$ are $1/N_0 + 1/N$. We wish to allocate fish to treatment and control groups so as to minimize $1/N_0 + 1/N$ subject to $N_0 + IN = C$, fixed. This is a Lagrange multiplier problem whose solution is $N_0 = NI^{1/2}$. Thus the more treatment groups, the greater is the sample size in the control group relative to that in the treatment groups. This is because the control group enters into all pairwise comparisons whereas each treatment group enters into just one. The suggested sample sizes are then

$$N = C/(I + I^{1/2})$$

$$N_0 = CI^{1/2}(I + I^{1/2})$$

This implies that for every 100 fish tested, the allocation between treatment and control groups would be

<u>I</u>	<u>Control (per 100 fish)</u>	<u>Treatment(per 100 fish)</u>
1	50	50
2	41.4	29.3
3	36.60	21.13
4	33.33	16.67
5	30.90	13.82
6	28.99	11.84
7	27.43	10.36
8	26.12	9.23
9	25	8.33
10	24.03	7.60
11	23.16	6.98
12	22.40	6.47

We see that the allocation is far from equal if I is moderate. For example if $I = 5$, the control group gets 2.5 times as many fish as any of the treatment groups.

How effective in increasing sensitivity of inferences is this departure from equal allocation? To determine the sensitivity of various sizes

of tests to detect increases in response rates between control group and treatment groups, we carried out a series of power calculations. The null and alternative hypotheses considered were:

$$H_0: p = p_0$$

$$H_1: p > p_0$$

We estimate p_0 , p by the sample response rates \hat{p}_0 , \hat{p} and reject H_0 at $\alpha = 0.05$ if

$$2 \arcsin \sqrt{\hat{p}} - 2 \arcsin \sqrt{\hat{p}_0} > 1.645(1/N_0 + 1/N)^{1/2}$$

simultaneity considerations are ignored in this calculation. The power of this test is calculated for various levels of N , p , p_0 . The expression for the power is

$$1 - \Phi \left[1.645 - \frac{(\phi_1 - \phi_0)}{\sqrt{1/N_0 + 1/N}} \right]$$

where $\Phi(\cdot)$ is the standard normal c.d.f.

$$\phi_1 = 2 \arcsin \sqrt{\hat{p}_1}$$

$$\phi_0 = 2 \arcsin \sqrt{\hat{p}_0}$$

Calculations were made for the cases of equal allocation (i.e. N fish per group) and "optimal" allocation (i.e. $N(I+1)/(I+I^{1/2})$ fish in each treatment group and $N(I+1)I^{1/2}/(I+I^{1/2})$ fish in the control group. The usual situation, $I = 5$, is considered. The results are shown in Table XVIII.1.

TABLE XVIII.1 POWER OF ONE SIDED PAIRWISE COMPARISONS OF SURVIVAL RATES
BETWEEN CONTROL GROUP AND TREATMENT GROUPS. $\alpha = 0.05$.

<u>$p_o = 0.001$</u>	<u>N</u>	<u>50</u>	<u>75</u>	<u>100</u>	<u>150</u>	<u>200</u>
p = .05		0.62*	0.77	0.87	0.96	0.988
		0.67*	0.82	0.90	0.98	0.995
.10		0.87	0.96	0.99	0.999	1.000
		0.91	0.98	0.995	1.000	1.000
.15		0.96	0.995	0.999		
		0.98	0.998	1.000		
.20		0.99	1.000			
		0.996	1.000			
.30		1.000				
		1.000				
<u>$p_o = .05$</u>	<u>N</u>	<u>50</u>	<u>75</u>	<u>100</u>	<u>150</u>	<u>200</u>
p = .10		0.25	0.32	0.39	0.51	0.61
		0.27	0.35	0.42	0.55	0.66
.15		0.53	0.68	0.79	0.91	0.96
		0.58	0.73	0.83	0.94	0.98
.20		0.77	0.90	0.96	0.993	0.999
		0.82	0.93	0.98	0.997	1.000
.30		0.97	0.996	1.000	1.000	
		0.98	0.999	1.000	1.000	
.40		0.998	1.000			
		1.000	1.000			

*Top number represents power under equal allocation. Bottom number represents power under optimal allocation, I = 5.

<u>$p_o = .10$</u>	<u>N</u>	<u>50</u>	<u>75</u>	<u>100</u>	<u>150</u>	<u>200</u>
$p = .15$		0.19	0.24	0.28	0.37	0.45
		0.20	0.26	0.31	0.41	0.49
.20		0.41	0.54	0.64	0.79	0.88
		0.45	0.58	0.69	0.84	0.92
.30		0.83	0.93	0.98	0.998	1.000
		0.87	0.96	0.99	0.999	1.000
.40		0.98	0.997	1.000	1.000	
		0.99	0.999	1.000	1.000	
.50		0.999	1.000			
		1.000	1.000			

<u>$p_o = 0.15$</u>	<u>N</u>	<u>50</u>	<u>75</u>	<u>100</u>	<u>150</u>	<u>200</u>
$p = .20$		0.16	0.20	0.24	0.31	0.37
		0.18	0.22	0.26	0.34	0.41
.30		0.57	0.72	0.82	0.93	0.98
		0.62	0.77	0.86	0.96	0.99
.40		0.89	0.97	0.992	1.000	1.000
		0.92	0.98	0.997	1.000	1.000
.50		0.99	0.999	1.000		
		0.994	1.000	1.000		

The following conclusions can be drawn from Table XVIII.1.

1. The ability to discriminate between treatment group and control group mortality rates varies considerably with N , p_0 , p . Thus these calculations provide some idea of the discrimination capability of the test as a function of size. Remember of course that these calculations don't account for the effects of heterogeneity among tanks.
2. The effect of allocating more fish to the control group than to the treatment groups is minor. Equal allocation yields nearly as good power as "optimal" allocation and is logistically much simpler.
3. If no assumptions can be made about the magnitudes of survival rates to be expected at the various concentration groups then equal allocations should be used.
4. If we can say something a priori about the survival rates at the various treatment groups, then we should have larger sample sizes in the lower concentration groups and smaller sample sizes in the higher concentration groups. The aim should be to even out the power of treatment group-control group comparisons as much as possible across groups.

We now consider adjustments in power calculations to take account of tank to tank heterogeneity. We do this by adjusting downward the effective sample size in each group and then entering Table XVIII.1 with the effective sample size. We calculate the adjustment by use of the beta binomial model [21].

Suppose that the test consists of I groups, J tanks per group, and n organisms (fish or embryos) per tank. Thus the actual sample size per group is $N \equiv Jn$. Let X_{ij} denote the number of responses (dead, abnormal, etc) in the j -th tank of group i . We assume that X_{ij} is binomially distributed with parameters (n, p_{ij}) and p_{ij} is in turn beta distributed with parameters (α_i, β_i) . This model allows for random variation of the p_{ij} 's within the i -th group. Let

$$\mu_i \equiv \alpha_i / (\alpha_i + \beta_i)$$

$$\theta_i \equiv 1 / (\alpha_i + \beta_i)$$

$$\text{Then } E(p_{ij}) = \mu_i \qquad \text{Var}(p_{ij}) = \mu_i(1 - \mu_i)\theta_i / (1 + \theta_i)$$

The unconditional distribution of X_{ij} is beta binomial with mean and variance

$$E(X_{ij}) = n\mu_i$$

$$\text{Var}(X_{ij}) = n\mu_i(1 - \mu_i) \frac{1 + n\theta_i}{1 + \theta_i} \quad 0 \leq \theta_i < \infty$$

Assume that $\theta_i \equiv \theta$ is constant across groups. Let

$$K \equiv \frac{1 + n\theta}{1 + \theta}$$

This is the variance inflation factor due to tank to tank heterogeneity. The effective sample size per tank is then n/K and so the effective sample size per group is

$$N_{\text{eff}} = Jn/K = N \frac{1 + \theta}{1 + n\theta}$$

As $\theta \rightarrow 0$, $N_{\text{eff}} \rightarrow N$, the number of organisms. As $\theta \rightarrow \infty$, $N_{\text{eff}} \rightarrow J$, the number of tanks. As $\sigma \rightarrow \infty$, $N_{\text{eff}} \rightarrow J(1 + \theta)/\theta$. Thus the effective number of organisms per tank asymptotes out as the actual number increases. Figure XVIII.1 shows a plot of $n_{\text{eff}} \equiv n(1 + \theta)/(1 + n\theta)$ vs n for various values of θ . We see the diminishing returns of placing more and more fish per tank in the presence of tank to tank heterogeneity. However under the cost structure assumed in this section we still place the maximum number of organisms within each tank, which we assume is 50 for embryos and 25 for fry. These numbers of course are only working assumptions.

To get some feeling for the meaning of θ in terms of variance inflation factors, we calculate the factors corresponding to $n = 25$, $n = 50$ for various values of θ .

θ	0	0.01	0.025	0.05	0.10	0.50	1.00
Var.infl.fact., $n=25$	1	1.24	1.61	2.19	3.18	9	13
Var.infl.fact., $n=50$	1	1.5	2.2	3.3	5.5	17.3	25.5

We calculated variance inflation factors for several sets of Fathead Minnow data in Section IX. These were

1. Holcombe and Phipps compound D fry mortality

$$n = 25, \quad \hat{K} = 1.337 \equiv (1 + 25\theta)/(1 + \theta)$$

Thus $\hat{\theta} = 0.014$ and $n_{\text{eff}} = 18.78$. Thus $N_{\text{eff}} = 75$.

2. Jarvinen compound B embryo mortality

$$n = 50, \quad \hat{K} = 3.071 \equiv (1 + 50\theta)/(1 + \theta)$$

Thus $\hat{\theta} = 0.044$ and $n_{\text{eff}} = 16.31$. Thus $N_{\text{eff}} = 65$.

C. Expected Precision for Comparison of Treatment Group and Control Group Survival Probabilities

In the previous subsection we calculated the power to be expected for pairwise treatment group-control group comparisons of survival probabilities as a function of p_0 , p , N . In this subsection we calculate the expected half lengths of two sided 95 percent confidence intervals on $p - p_0$ for these same combinations of p_0 , p , N . We again assume no tank to tank heterogeneity within groups and account for such heterogeneity by reducing the effective sample sizes, as discussed in the previous subsection. We base the precision calculations on asymptotic normal theory. Namely the confidence interval half length is calculated as $1.96[(p_0 q_0 + pq)/N]^{1/2}$. Asymptotic normality may not be a very good assumption when $N = 50$ or when $p_0 = 0.001$, but it is only being used for planning purposes.

TABLE XVIII.2 EXPECTED HALF LENGTHS OF 95 PERCENT TWO SIDED CONFIDENCE INTERVALS FOR COMPARISONS BETWEEN TREATMENT GROUP AND CONTROL RESPONSE RATES

$p_0 = 0.001$	N	50	75	100	150	200
$p = .05$.06	.05	.04	.04	.03
.10		.08	.07	.06	.05	.04
.15		.10	.08	.07	.06	.05
.20		.11	.09	.08	.06	.06
.30		.13	.10	.09	.07	.06
.40		.14	.11	.10	.08	.07
.50		.14	.11	.10	.08	.07
$p_0 = 0.05$						
$p = .10$.10	.08	.07	.06	.05
.15		.12	.09	.08	.07	.06
.20		.13	.10	.09	.07	.06
.30		.14	.11	.10	.08	.07
.40		.15	.12	.10	.09	.07
.50		.15	.12	.11	.09	.08
$p_0 = 0.10$						
$p = .15$.13	.11	.09	.07	.06
.20		.14	.11	.10	.08	.07
.30		.15	.12	.11	.09	.08
.40		.16	.13	.11	.09	.08
.50		.16	.13	.11	.09	.08

<u>$p_o = 0.15$</u>	<u>N</u>	<u>50</u>	<u>75</u>	<u>100</u>	<u>150</u>	<u>200</u>
p = .20		.15	.12	.11	.09	.07
.30		.16	.13	.11	.09	.08
.40		.17	.14	.12	.10	.08
.50		.17	.14	.12	.10	.09

D. Power Calculations for Quantitative Weight Response

In previous subsections we calculated expected power and expected estimation precision for the quantal survival response. In this subsection we carry out similar calculations for the quantitative weight response. Distributional assumptions are based on the results of analyzing the Holcombe and Phipps compound D fry weights in Section XVII. There was considerable group to group variation in survival proportions but not as much variation in the weights of the surviving fry. In particular

	<u>Control Group</u>	<u>Group 4</u>	<u>Group 5</u>	<u>Group 6</u>
Survivors	94/100	87/100	21/100	0/100
Avg. Wt. of Survivors (mg)	131.2	113.1	108	
Max. Wt.	190	186	195	
Min. Wt.	45	29	34	
Std. Dev.	25.8	34.0	36.7	

Variability is not too dependent on concentration group or on survival rate. The variance components are assumed to be constant across treatment groups.

The power and precision calculations below are based on a number of assumptions.

1. There is no tank to tank variation within treatment groups. We discuss corrections for such factors later in the subsection.
2. Equal sample sizes among treatment and control groups.

This assumption is reasonable if we confine comparisons of weight gains to treatment groups with mortality rates not greatly in excess of the control rate. Otherwise an average or minimum N might be used.

3. Constant variability across treatment groups.

This assumption might hold for the weights themselves or for some function of the weights such as log weights.

4. There are enough observations to have effectively an infinite number of degrees of freedom. The power obtained with 12 d.f. is nearly that obtained with infinite d.f.

5. No simultaneity correction is applied.

Table XVIII.3 shows the power of a one sided normal theory test of

$$\begin{array}{l} H_0: \mu = \mu_0 \\ \text{vs} \\ H_1: \mu < \mu_0 \end{array}$$

where μ , μ_0 are the average weights in the treatment and control groups respectively. In the absence of tank effects the individual weights are assumed to have standard deviation σ . The bottom portion of the table contains factors C , for constructing 95 percent lower confidence bounds on $\mu_0 - \mu$. Namely $\mu_0 - \mu \geq \bar{X}_0 - \bar{X} - C\sigma$.

TABLE XVIII.3 POWER OF ONE SIDED PAIRWISE COMPARISONS OF AVERAGE WEIGHT GAINS BETWEEN CONTROL GROUP AND TREATMENT GROUPS. $\alpha = 0.05$

$d = (\mu_o - \mu)/\sigma$	N	10	15	25	35	50	75	100	150	200
.02	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07
.04	0.06	0.06	0.07	0.07	0.07	0.07	0.08	0.09	0.10	0.11
.06	0.07	0.07	0.08	0.08	0.08	0.09	0.10	0.11	0.13	0.15
.08	0.07	0.08	0.09	0.10	0.11	0.11	0.12	0.14	0.17	0.20
.1	0.08	0.09	0.10	0.11	0.13	0.13	0.15	0.17	0.22	0.26
.2	0.12	0.14	0.18	0.21	0.26	0.26	0.33	0.41	0.54	0.64
.3	0.17	0.21	0.28	0.35	0.44	0.44	0.58	0.68	0.83	0.91
.4	0.23	0.29	0.41	0.51	0.64	0.64	0.79	0.88	0.97	0.99
.5	0.30	0.39	0.55	0.67	0.80	0.80	0.92	0.97	0.996	1.000
.6	0.38	0.50	0.68	0.83	0.91	0.91	0.98	0.995	1.000	
.7	0.47	0.61	0.80	0.90	0.97	0.97	0.996	1.000		
.8	0.56	0.71	0.88	0.96	0.991	0.991	0.999			
.9	0.64	0.79	0.94	0.98	0.998	0.998	1.000			
1	0.72	0.86	0.97	0.995	1.000	1.000				
1.2	0.85	0.95	0.995	1.000						
1.4	0.93	0.99	1.000							
1.6	0.97	0.997								
1.8	0.991	1.000								
2	0.998									
2.2	1.000									
C	.74	.60	.47	.39	.33	.27	.23	.19	.16	

The power calculations and precision factors in Table XVIII.3 need to be adjusted for tank to tank variation within groups. Suppose that there are J tanks per group, n organisms per tank, and that $\sigma_{\tau}^2, \sigma_e^2$ represent the between and within tank components of variation. The variance of the average weight in the group is then

$$\frac{\sigma_{\tau}^2}{J} + \frac{\sigma_e^2}{N} = \frac{\sigma_e^2 + n\sigma_{\tau}^2}{N}$$

Table XVIII.3 is entered at N and at $d = (\mu_o - \mu) / (\sigma_e^2 + n\sigma_{\tau}^2)^{1/2}$

The precision factors at the bottom of the table remain the same but the confidence bound is $\mu_o - \mu \geq \bar{X}_o - \bar{X} - C(\sigma_e^2 + n\sigma_{\tau}^2)^{1/2}$

We now apply these relations to the weight gain data from the Holcombe and Phipps test on compound D. In that example there is no statistical evidence of tank to tank variation within groups. Namely

$$\hat{\sigma}_e^2 = 900.354 \quad \text{with 366 d.f.}$$

$$\hat{\sigma}_e^2 + 19.191\hat{\sigma}_{\tau}^2 = 754.97 \quad \text{with 15 d.f.}$$

Thus

$$\hat{\sigma}_{\tau}^2 = -7.58$$

and we assume it is 0. Since σ_e^2 is estimated with 366 degrees of freedom, we assume it is known exactly. Thus $\hat{\sigma}_e = 30.0$. The average N in groups 1-4 is 91.24 while the sample size in group 5 is 21. Assume for the purpose of power calculations that $\mu_o = \bar{X}_o, \mu = \bar{X}$. Thus $\mu_o = 131.245, \mu_3 = 127.924, \mu_4 = 113.080, \mu_5 = 108.095$.

Therefore,

$$d_3 \equiv \frac{\mu_o - \mu_3}{\sigma} = \frac{131.245 - 127.924}{30.0} = 0.111$$

$$d_4 \equiv \frac{\mu_o - \mu_4}{\sigma} = \frac{131.245 - 113.080}{30} = 0.606$$

$$d_5 \equiv \frac{\mu_o - \mu_5}{\sigma} = \frac{131.245 - 108.095}{30} = 0.772$$

Interpolating (approximately) in Table XVIII.3 with $N = 90$ and d_3 , d_4 , d_5 yields

Control vs group 3 Power = 0.16

Control vs group 4 Power = 0.99

For group 5 the assumption of equal N is not reasonable and so we calculate the noncentrality parameter for the test as $d_5/(1/N_0 + 1/N_5)^{1/2} = 3.186$. Thus power = $\Phi(\text{noncentrality} - 1.645) = \Phi(3.186 - 1.645) = \Phi(1.540) = 0.94$.

E. Unequal Allocations of Testing Effort Among Treatment Groups

Standard test guidelines call for equal numbers of tanks to be run at each treatment group. Such a design would be sensible only if prior to running the test there was total ignorance about response levels to be expected. That is suppose it was thought a priori that the mortality rate for each treatment group could be anywhere between 0 and 100 percent. Then it would make good sense to allocate experimental effort equally among treatment groups to assure specified power whenever and wherever the mortality rate exceeds that in the control group by a specified amount. However if on the basis of either a priori scientific information or previous testing some information was available concerning mortality rates to be expected at the various treatment groups, then unequal allocation of experimental effort would be preferable. In particular at the higher treatment groups, where mortality would be expected to be substantially higher than the control rate, it is easy to detect differences from the control. Thus the experimental effort should be decreased at these groups. At the lower experimental groups, where it is more difficult to detect differences from the control group, the experimental effort should be increased to improve sensitivity. Thus the degree of experimental effort should in general decrease as the toxicant level increases.

Details of a procedure for arriving at an unequal allocation will be discussed in the report on phase 2, for Daphnia magna. For the purpose of this subsection, consider the following illustration of the effects on sensitivity of unequal replication. Suppose that the experiment is to consist of a control group and $I = 5$ treatment groups. Suppose that cost and logistical restraints limit the number of tanks to 24, that $n = 25$ fry will be exposed in each tank and that tank to tank heterogeneity is such that the variance inflation factor is 1.5. Suppose it is felt that the control group response rate will be about 0.05 and the mortality rates in the treatment groups will be about 0.10, 0.15, 0.20, 0.40, and 0.80 respectively. The classical allocation would be to run $J = 4$ tanks per group. Thus $N = 100$ fish per group. The effective sample sizes would be $N_{\text{eff}} = 100/1.5 = 66.67$ fish per group. Suppose further that it is considered important (biologically and/or legally) to detect increases in mortality of 10 percent above the control rate. That is, we wish to detect differences between 5 percent and 15 percent mortality.

Under the classical allocation scheme the power to be expected for each treatment group-control group comparison would be:

Group 2 vs control

$$1 - \beta = \Phi[(2 \arcsin \sqrt{.10} - 2 \arcsin \sqrt{.05}) / (2/66.67)^{1/2} - 1.645] = \\ \Phi[(.64 - .45) / (2/66.67)^{1/2} - 1.645] = \Phi(-0.53) = 0.30$$

Group 3 vs control

$$1 - \beta = \Phi[(.80 - .45)/(2/66.67)^{1/2} - 1.645] = \Phi(0.38) = 0.65$$

Group 4 vs control

$$1 - \beta = \Phi[(.93 - .45)/(2/66.67)^{1/2} - 1.645] = \Phi(1.13) = 0.87$$

Group 5 vs control

$$1 - \beta = \Phi[(1.37 - .45)/(2/66.67)^{1/2} - 1.645] = \Phi(3.67) = 1.000$$

Group 6 vs control

$$1 - \beta = \Phi[(2.21 - .45)/(2/66.67)^{1/2} - 1.645] = \Phi(8.52) = 1.000$$

Consider the modified allocation of 7, 6, 6, 3, 1, 1 tanks in the control group and in each of the treatment groups respectively. Then $N_0 = 175$, $N_2 = 150$, $N_3 = 150$, $N_4 = 75$, $N_5 = N_6 = 25$. The effective sample sizes, $N_{eff} = N/1.5$, are then 116.67, 100, 100, 50, 16.67, 16.67. The power to be expected for each treatment group-control group comparison would then be:

Group 2 vs control

$$1 - \beta = \Phi[(2 \arcsin \sqrt{.10} - 2 \arcsin \sqrt{.05})/(1/116.67 + 1/100)^{1/2} - 1.645] = \Phi[(.64 - .45)/0.136 - 1.645] = \Phi(-0.25) = 0.40$$

Group 3 vs control

$$1 - \beta = \Phi[(.80 - .45)/(1/116.67 + 1/100)^{1/2} - 1.645] = \Phi(0.923) = 0.82$$

Group 4 vs control

$$1 - \beta = \Phi[(.93 - .45)/(1/116.67 + 1/50)^{1/2} - 1.645] = \Phi(1.195) = 0.88$$

Group 5 vs control

$$1 - \beta = \Phi[(1.37 - 0.45)/(1/116.67 + 1/16.67)^{1/2} - 1.645] = \Phi(1.87) = 0.97$$

Group 6 vs control

$$1 - \beta = \Phi[(2.21 - 0.45)/(1/116.67 + 1/16.67)^{1/2} - 1.645] = \Phi(5.08) = 1.00$$

Comparison of the two sets of calculations shows that we have improved the power of the comparisons at the low concentration end of the test without sacrificing any appreciable power at the high concentration end of the test. This has been done without increasing the size of the test. If we were willing to add several additional tanks we could do even better. The power for the comparison of the mortality rate 0.15 vs the control rate has been

increased from 0.65 to 0.82. This is a substantial improvement because the chance of not detecting such an increase in mortality has diminished from 1 in 2.86 to 1 in 5.56. This just about halves the type 2 error probability.

The previously suggested unequal allocation of experimental effort to concentration groups is intuitively sensible, improves sensitivity at the low end of the experiment where it is most needed, does not diminish sensitivity at the high end of the experiment, and does not increase the overall size of the test. It requires specifying prior beliefs about the response levels to be expected at the various concentration levels. A scheme for doing this will be discussed in the phase 2 report. However, it is pretty clear that the more definitive the prior information, the more unequal should the allocation be. As prior information diminishes to complete ignorance the design should approach equal allocation. Of course the efficacy of the design depends on the accuracy of the prior information. If it is believed a priori that group 6 will have between a 75 percent and 100 percent mortality rate but it in fact has a 15 percent mortality rate, then an unequal tank allocation scheme will probably do worse than an equal tank allocation scheme.

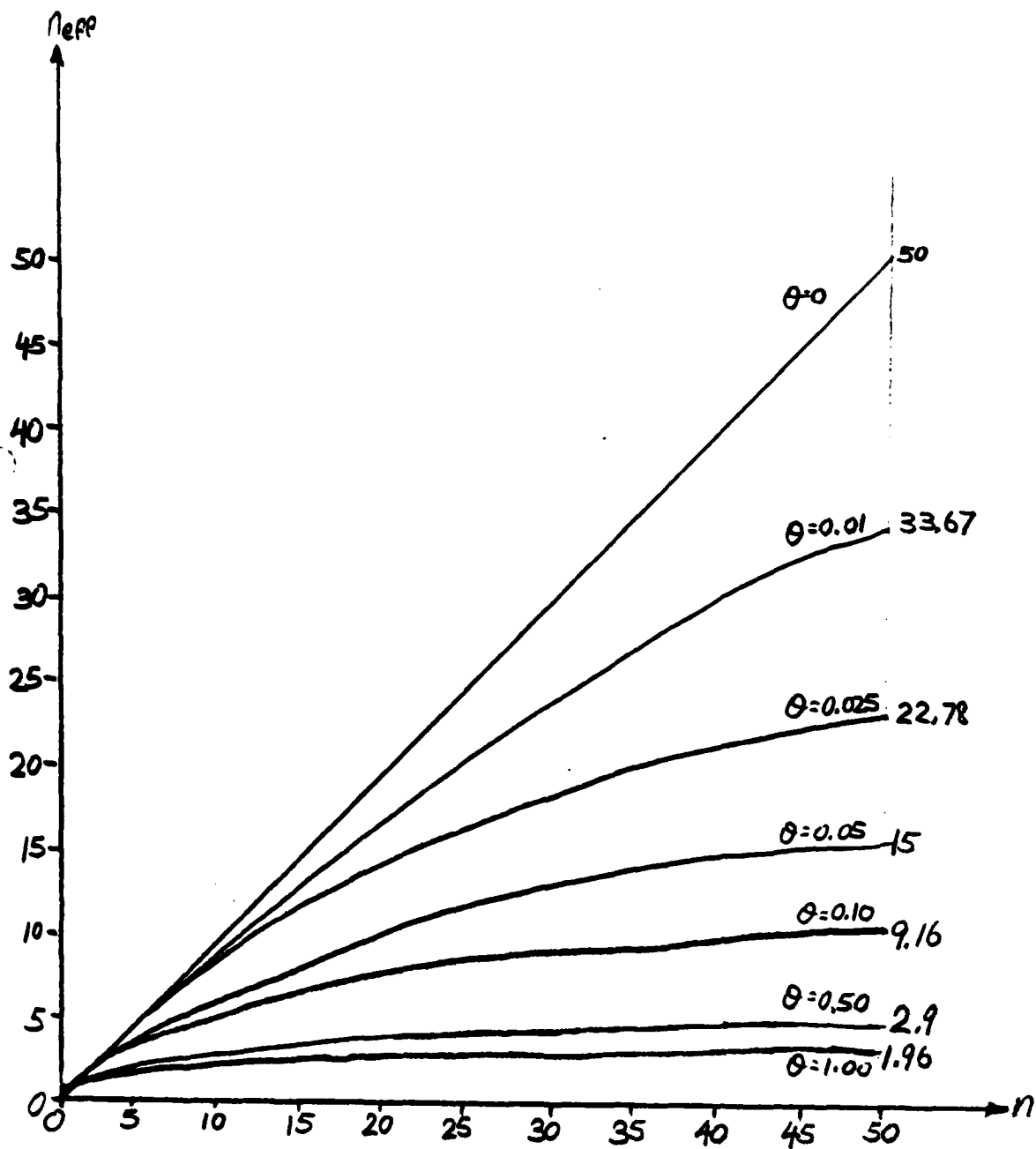


Figure XVIII.1 Effective sample size per group vs sample size for various values of θ

XIX DESIGN AND ANALYSIS CONSIDERATIONS FOR FULL LIFE CYCLE TESTS
DISTINCT FROM THOSE FOR EARLY LIFE STAGE TESTS

In recent years, efforts in aquatic toxicity testing have been shifting more and more to early life stage tests and away from full life cycle tests. The results obtained from a full life cycle test are directly analagous to those obtained from early life stage tests, only much more of them are accumulated. Mortality rates are recorded periodically, length and weight measurements are obtained periodically, and fecundity responses such as embryos per spawn, total numbers of embryos, spawns per female are recorded. With respect to the survival, weight, and length responses the design and analysis considerations discussed in the previous sections are directly applicable and require no amplifications or modifications.

Some differences in design and analysis considerations may be called for with respect to the fecundity responses. Questions of homogeneity of variances, tank to tank heterogeneity within groups, form of dose response relation are handled in essentially the same manner as the analagous questions for the mortality and weight responses. Similarly, sample size determination and tank allocation design calculation need to be made in the same manner as those carried out for mortality and for weight responses. The statistical issues are the same, but the numbers may turn out to be different.

One design consideration associated with fecundity responses may well introduce important differences as compared with those for mortality and weight considerations. Namely it has been assumed that the cost structure is such that there is a certain incremental cost associated with adding an additional tank to the test but once the tank is added, the fish are free. This leads to the recommendation to run as many tanks as can be afforded and fill each tank with the maximum number of embryos or fry that is biologically sensible. This cost structure may not hold for fecundity responses. There is a considerable amount of operational and clerical effort associated with accumulating the hatched embryos, counting them, associating them with the appropriate fish or groups of fish, and properly recording the data. The numbers of embryos produced are related to the numbers of fish rather than to the numbers of tanks. Thus fish can no longer be considered free. Effects of competition on production must also be considered. In the presence of tank effects it may thus be sensible to increase the number of tanks in the test and decrease the number of fish per tank. This may improve the precision of statistical inferences without incurring additional expense. The particular trade off between number of tanks and number of fish per tank would of course depend on the extent of tank to tank heterogeneity and on the tank and fish costs.

All of these issues arise in the design and analysis of toxicity tests on Daphnia magna. Survival, length, and fecundity responses are reported at periodic intervals. The fecundity responses reported by various investigators include average embryos per surviving female per chamber, embryos per surviving female, total number of embryos depending on the design of the test. Questions of multiple daphnids or individual daphnids

per beaker are commonly posed. Thus most of the design and analysis considerations involved in full life cycle tests that are distinct from those in the early life stage tests will be addressed in the course of the discussion of the design and analysis of the Daphnia toxicity tests.

REFERENCES

- [1] Manual for Construction and Operation of Toxicity - Testing Proportional Diluters. USEPA Environmental Research Laboratory- Duluth, Minnesota, July, 1978 (EPA 600/3-78-072).
- [2] "Recommended Bioassay Procedure for Fathead Minnow Pimephales Promelas Rafinesque Chronic Tests," USEPA Environmental Research Laboratory, Duluth, Minnesota, January, 1972.
- [3] DeFoe, D. L., "Multichannel Toxicant Injection System For Flowthrough Bioassays," Journal Fisheries Research Board of Canada, Vol. 32, 1975, pp. 544-546.
- [4] Federal Register, Vol. 44, No. 145, July 24, 1979, pp. 44054-44093.
- [5] Fishbeck, R. D., American Laboratory, September, 1980, pp. 125-130.
- [6] F. McKim, J. M., "Evaluation of Tests With Early Life Stages of Fish for Predicting Long Term Toxicity, Journal Fisheries Research Board of Canada, Vol. 34, 1977, pp. 1148-1154.
- [7] Federal Register, Vol. 45, No. 98, May 19, 1980, pp. 33124-127.
- [8] Stephan, C. et al, "Guidelines for Conducting Flow-Through Early Life Stage Toxicity Tests with Fathead Minnows for use in the USEPA, OTS-ORD Round Robin Test," USEPA Environmental Research Laboratory, Duluth, Minnesota, November, 1979.
- [9] Stephan, C. et al, "Early Life Stage Toxicity Tests with Fathead Minnows," USEPA Environmental Research Laboratory, Duluth, Minnesota memorandum, October, 1979.
- [10] "Standard Practice for Conducting Toxicity Tests on the Early Life Stages of Fishes," Working Draft of ASTM Committee E-35 on Pesticides, September, 1979.
- [11] Finney, D. J., Probit Analysis, Third Edition, Cambridge University Press, London, 1971, pp. 175 ff., page 125.
- [12] Helwig, J. T. and Council, K. A., Editors, SAS User's Guide, 1979 Edition, SAS Institute, Inc. Raleigh, 1979.
- [13] Dixon, W. J. and Massey, F. J., Jr., Introduction to Statistical Analysis, McGraw-Hill, New York, 1969, pp. 238, 240-243.
- [14] Feder, P. I. and Willavize, Susan A., "EXAX2 -- A Computer Program to Compare Binomial Proportions, Program Description and Card Input Information," Technical Report No. 216, Department of Statistics, The Ohio State University, Columbus, Ohio, September 1980.

- [15] Agresti, A. and Wackerly, D., "Some Exact Conditional Tests of Independence for RxC Crossclassification Tables," Psychometrika, Vol. 42, No. 1, March, 1977, pp. 111-125.
- [16] Boulton, D. M. and Wallace, C. S., "Occupancy of a Rectangular Array," Computer Journal, Vol. 16, 1973, pp. 57-63.
- [17] Boulton, D. M., "Remark on Algorithm 434," Communications of the ACM, Vol. 17, 1974, pp. 326.
- [18] March, D. L., "Algorithm 434, Exact Probabilities for rxc Contingency Tables," Communications of ACM, Vol. 15, 1972, pp. 991-992
- [19] Littell, R. C. and Folks, J. L., "Asymptotic Optimality of Fishers Method of Combining Independent Tests," Journal of the American Statistical Association, Vol. 66, No. 336, December, 1971, pp. 802-806.
- [20] Littell, R. C. and Folks, J. L., "Asymptotic Optimality of Fisher's Method of Combining Independent Tests II," Journal of the American Statistical Association, Vol. 68, No. 341, March, 1973, pp. 193-194.
- [21] Williams, D. A., "The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity," Biometrics, Vol. 31, 1975, pp. 949-952.
- [22] Kupper, L. L. and Haseman, J. K., "On the Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments," Biometrics, Vol. 34, No. 1, 1978, pp. 69-76.
- [23] Barnett, V. and Lewis, F., Outliers in Statistical Data, Wiley, New York, 1978.
- [24] Nelson, W., "Confidence Intervals for the Ratio of Two Poisson Means and Poisson Predictor Intervals," General Electric Company Research and Development Center TIS Report 69C118, February, 1969. IEEE Transactions on Reliability, 1970, Vol. R-19, pp. 42-49.
- [25] Lehmann, E. L., Testing Statistical Hypotheses, 1959, John Wiley and Sons, Inc., Section 4.5.
- [26] Lieberman, G. J. and Owen D. B., Tables of the Hypergeometric Probability Distribution, 1961, Stanford University Press.
- [27] Dixon, W. J. and Brown, M. B., Editors, BMDP Biomedical Computer Programs P-Series 1979, University of California Press, Berkeley, 1979.
- [28] Snedecor, G. W. and Cochran, W. G., Statistical Methods, Sixth Edition, Iowa State University Press, Ames, 1967, p. 246.

- [29] Steel, R. G. D. and Torrie, J. H., Principles and Procedures of Statistics - A Biometrical Approach, Second Edition, McGraw-Hill, New York, 1980 p. 514.
- [30] Goodman, L. A. and Kruskal, W. H., "Measures of Association for Crossclassifications," Journal of the American Statistical Association, Vol. 49, December, 1954, pp. 732-764.
- [31] Goodman, L. A. and Kruskal, W. H., "Measures of Association for Crossclassifications, IV: Simplification of Asymptotic Variances," Journal of the American Statistical Association, Vol. 67, No. 338, June, 1972, pp. 415-421.
- [32] Brown, M. B., and Benedetti, J. K., "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables," Journal of the American Statistical Association, Vol. 72, No. 358, June, 1977, pp. 309-315.
- [33] Proctor, C. H., "Relative Efficiencies of Some Measures of Association for Ordered Two-Way Contingency Tables Under Varying Intervalness of Measurement Errors," Proceedings of the Social Statistics Section, American Statistical Association, 1973, pp. 372-379.
- [34] Dunnett, C. W., "A Multiple Comparisons Procedure for Comparing Several Treatments With a Control," Journal of the American Statistical Association, Vol. 50 pp. 1096-1121.
- [35] Dunnett, C. W., "New Tables for Multiple Comparisons With a Control," Biometrics, Vol. 20, 1964, pp. 482-491.
- [36] Williams, D. A., "A Test for Differences Between Treatment Means When Several Dose Levels are Compared With a Zero Dose Control," Biometrics, Vol. 27, 1971, pp. 103-117.
- [37] Williams, D. A., "The Comparison of Several Dose Levels With a Zero Dose Control," Biometrics, Vol. 28, 1972, pp. 519-531.
- [38] Chew, V., Comparisons Among Means in an Analysis of Variance, Agricultural Research Service Publication ARS/H/6, October, 1977.
- [39] Thomas, D. G., "Exact Confidence Limits for the Odds Ratio in a 2x2 Table," Applied Statistics, Vol. 20, 1971, pp. 105-110.
- [40] Box, G. E. P., Hunter, J. S., and Hunter, W. G., Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building, Wiley, New York, 1978.
- [41] Feder, P. I. and Sherrill, M. E., "A Computer Program to Calculate Nonparametric Lower Confidence Bounds on Safe Concentration in Quantal Response Toxicity Tests," Technical Report No. 215, Department of Statistics, The Ohio State University, Columbus, Ohio, September, 1980.

- [42] Stephan, C. E., "Methods for Calculating an LC₅₀," in Aquatic Toxicology and Hazard Evaluation, F. L. Mayer and J. L. Hamelink, Editors, ASTM Special Technical Publication 634, American Society for Testing and Materials, 1977, pp. 65-84.
- [43] Jennrich, R. I. and Moore, R. H., "Maximum Likelihood Estimation by Means of Nonlinear Least Squares," in Proceedings of the Statistical Computing Section, American Statistical Association, 1975, pp. 57-65
- [44] Mandel, John, The Statistical Analysis of Experimental Data, John Wiley and Sons, Inc., 1964.
- [45] Graybill, F. A., An Introduction to Linear Statistical Models, Volume I, McGraw-Hill Book Company, 1961, pp. 186-194.
- [46] Clopper, C. J. and Pearson, E. S., "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," Biometrics Vol. 26, 1934, pp. 404-413.
- [47] Hollander, M. and Wolfe, D. A., Nonparametric Statistical Methods, John Wiley and Sons, Inc., 1973.
- [48] Natrella, M. G., Experimental Statistics, National Bureau of Standards Statistics Handbook No. 91, U. S. Government Printing Office, 1963.
- [49] Gross, A. J. and Clark, V. A., Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley and Sons, Inc., 1975.
- [50] Kalbfleisch, J. D. and Prentice, R. L., The Statistical Analysis of Failure Time Data, John Wiley and Sons, Inc., 1980.

APPENDIX AII* EARLY LIFE STAGE DATA SETS USED AS EXAMPLES IN THE BODY OF THE REPORT

This appendix contains listings of data sets from four early life stage toxicity tests. These data are used for illustrations of the procedures discussed in the body of the report. These data sets are

Benoit - compound A
DeFoe - compound C
Holcombe and phipps - compound D
Jarvinen - compound B.

The three types of data -- survival, weight and toxicant concentration -- are represented in three "card types." The first six entries on each card are the same across card types -- treatment group (col 2), replicate designation (col 4), card type (col 6), card member (cols 7=8), investigator code (cols 9-10), test code (cols 11-12). This provides enough information to sort the cards by investigator, experiment, type, group, and sequence should the data become disarranged. Card type 1 (survival data) contains in addition number of embryos tested (cols 16-20), number hatched live (cols 21-25), number of fry tested (cols 31-35), number live at end of test (cols 36-40), number normal at end of test (cols 41-45). Card type 2 (weight data) contains number of weights recorded from that particular chamber (cols 14-15), individual weights (5 cols per weight, up to 13 weights per card). Card type 3 (toxicant concentration) contains month (cols 16-17), day (cols 18-19), year (cols 20-21), toxicant concentration (cols 32-38) -- one determination per card. At the head of each type of information several lines of descriptive text are given.

* Appendix AII is the appendix for Section II.

VESTIGATOR: JARVINON (04), TEST: COMPOUND B

(01)

TA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS

X LEVELS: 1(CONTROL),2(LOWEST),...,6(HIGHEST); 2 REPS EA (A,B)

EMBRYOS TESTED, # LIVE AFTER HATCH, # NORMAL FRY AFTER HATCH

FRY TESTED, # ALIVE AT END, # NORMAL AT END

A	1	10401	51	39	39	16	16	16
B	1	10401	50	48	48	14	14	14
A	1	10401	57	51	51	15	15	15
B	1	10401	52	50	50	14	14	14
A	1	10401	50	48	48	15	15	15
B	1	10401	50	43	43	15	15	15
A	1	10401	50	49	49	15	15	15
B	1	10401	48	47	47	15	15	15
A	1	10401	52	48	48	15	08	08
B	1	10401	53	51	51	15	06	06
A	1	10401	50	49	49	15	00	00
B	1	10401	51	44	44	15	00	00

VESTIGATOR: JARVINON (04), TEST: COMPOUND B

(01)

TA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS

X LEVELS: 1(CONTROL),2(LOWEST),...,6(HIGHEST); 2 REPS EA (A,B)

DIVIDUAL WEIGHTS (MG) OF ALL FISH ALIVE AT END OF TEST (1-2 CARDS/TANK)

MBER OF WEIGHTS & LIST OF WEIGHTS

A	2	10401	16	096	080	072	098	059	074	111	078	104	097	060	057	091
A	2	20401	16	085	070	064										
B	2	10401	14	107	090	069	082	073	069	120	078	092	120	099	135	069
B	2	20401	14	068												
A	2	10401	15	093	096	096	093	100	083	098	043	090	070	069	040	052
A	2	20401	15	072	068											
B	2	10401	14	114	085	079	112	070	082	099	090	084	078	100	061	061
B	2	20401	14	071												
A	2	10401	15	080	100	046	100	105	082	071	066	096	078	087	078	068
A	2	20401	15	081	067											
B	2	10401	15	088	090	110	078	087	070	041	059	117	064	039	047	034
B	2	20401	15	053	064											
A	2	10401	15	093	091	089	091	058	086	058	070	062	048	083	048	065
A	2	20401	15	098	075											
B	2	10401	15	089	077	084	078	060	075	090	056	072	065	073	065	050
B	2	20401	15	083	064											
A	2	10401	08	078	060	052	065	067	047	045	105					
B	2	10401	06	078	088	070	038	102	048							
A	2	10401	00													
B	2	10401	00													

INVESTIGATOR: JARVINON (04), TEST: COMPOUND B

(01)

DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 2 REPS EA (A,B)
MEASURED CONCENTRATIONS OF TOXICANT (MG/L)

1 A 3 10401	030879	0.00
1 A 3 20401	031579	0.00
1 A 3 30401	032279	0.00
1 A 3 40401	032979	0.00
1 B 3 10401	030679	0.00
1 B 3 20401	031379	0.00
1 B 3 30401	032079	0.00
1 B 3 40401	032779	0.00
<hr/>		
2 A 3 10401	030879	0.22
2 A 3 20401	031579	0.24
2 A 3 30401	032979	0.19
2 B 3 10401	030679	0.29
2 B 3 20401	031379	0.23
2 B 3 30401	032079	0.28
2 B 3 40401	032779	0.18
3 A 3 10401	030879	0.34
3 A 3 20401	031579	0.39
3 A 3 30401	032279	0.53
3 A 3 40401	032979	0.25
3 B 3 10401	030679	0.48
3 B 3 20401	031379	0.36
3 B 3 30401	032079	0.42
3 B 3 40401	032779	0.29
4 A 3 10401	030879	0.46
4 A 3 20401	031579	0.55
4 A 3 30401	032279	0.89
4 A 3 40401	032979	0.33
4 B 3 10401	030679	0.72
4 B 3 20401	031379	0.52
4 B 3 30401	032079	0.88
4 B 3 40401	032779	0.35
5 A 3 10401	030879	0.60
5 A 3 20401	031579	0.73
5 A 3 30401	032279	0.91
5 A 3 40401	032979	0.61
5 B 3 10401	030679	1.10
5 B 3 20401	031379	0.59
5 B 3 30401	032079	1.10
5 B 3 40401	032779	0.51
6 A 3 10401	030879	1.10
6 A 3 20401	031579	0.96
6 A 3 30401	032279	1.90
6 A 3 40401	032979	1.10
6 B 3 10401	030679	1.40
6 B 3 20401	031379	0.69
6 B 3 30401	032079	1.60
6 B 3 40401	032779	1.10

INVESTIGATOR: DUANE BENOIT (01), TEST: A (01)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 4 REPS EA (A, B, C, D)
 # EMBRYOS TESTED, # ALIVE AFTER HATCH, # NORMAL FRY AFTER HATCH
 # FRY TESTED, # ALIVE AT END, # NORMAL AT END

1 A 1 10101	30	24	23	15	15	15
1 B 1 10101	30	26	26	15	15	15
1 C 1 10101	30	20	20	15	15	15
1 D 1 10101	30	24	23	15	15	15
2 A 1 10101	30	22	21	15	15	15
2 B 1 10101	30	21	21	15	14	14
2 C 1 10101	30	21	21	15	15	15
2 D 1 10101	30	16	16	15	15	15
3 A 1 10101	30	17	16	15	14	14
3 B 1 10101	30	20	20	15	14	14
3 C 1 10101	30	22	21	15	15	15
3 D 1 10101	30	22	21	15	15	15
4 A 1 10101	30	12	11	15	12	12
4 B 1 10101	30	18	18	15	11	11
4 C 1 10101	30	22	22	15	14	14
4 D 1 10101	30	25	24	15	14	14
5 A 1 10101	30	20	20	15	09	09
5 B 1 10101	30	17	17	15	06	06
5 C 1 10101	30	17	16	15	09	09
5 D 1 10101	30	24	24	15	06	08
6 A 1 10101	30	19	18	15	08	08
6 B 1 10101	30	20	19	15	03	03
6 C 1 10101	30	26	25	15	12	12
6 D 1 10101	30	21	21	15	10	10

INVESTIGATOR: DUANE BENOIT (01), TEST: A (01)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 4 REPS EA (A, B, C, D)
 INDIVIDUAL WEIGHTS (MG) OF ALL FISH ALIVE AT END OF TEST (1-2 CARDS/TANK)
 NUMBER OF WEIGHTS & LIST OF WEIGHTS

1 A 2 10101 14	164	152	164	123	130	094	141	150	205	128	086	070	116
1 A 2 20101 14	139												
1 B 2 10101 15	090	125	180	129	102	165	103	135	170	180	140	092	162
1 B 2 20101 15	139	106											
1 C 2 10101 15	175	130	102	143	121	131	172	120	150	090	133	125	121
1 C 2 20101 15	152	118											
1 D 2 10101 15	136	080	123	116	133	100	100	090	183	190	073	084	090
1 D 2 20101 15	152	160											
2 A 2 10101 15	090	240	125	082	084	120	190	114	120	108	107	158	128
2 A 2 20101 15	110	060											
2 B 2 10101 14	123	145	112	180	083	124	100	155	111	107	090	161	142
2 B 2 20101 14	094												
2 C 2 10101 15	083	134	114	118	180	160	130	180	132	190	084	120	132
2 C 2 20101 15	131	109											
2 D 2 10101 15	180	082	152	180	173	138	130	092	133	130	086	140	115
2 D 2 20101 15	093	137											
3 A 2 10101 14	110	159	094	148	080	151	121	096	144	132	100	112	095
3 A 2 20101 14	160												
3 B 2 10101 14	113	162	070	141	134	090	124	248	182	138	094	098	131
3 B 2 20101 14	146												
3 C 2 10101 15	110	073	240	119	130	110	090	084	080	193	123	086	150
3 C 2 20101 15	162	118											
3 D 2 10101 15	110	150	200	078	093	147	115	122	101	126	092	153	103
3 D 2 20101 15	120	070											
4 A 2 10101 12	131	130	112	132	092	132	112	099	106	140	125	094	
4 B 2 10101 11	083	136	148	060	162	139	157	121	091	146	134		
4 C 2 10101 15	182	086	140	106	105	165	130	184	093	115	122	099	145
4 C 2 20101 15	117	094											

4 D	2	10101	14	121	140	250	132	132	114	123	105	116	154	120	131	072
4 D	2	20101	14	111												
5 A	2	10101	09	043	111	100	064	070	111	090	050	150				
5 B	2	10101	06	054	117	127	098	121	157							
5 C	2	10101	09	070	065	184	144	125	084	112	071	106				
5 U	2	10101	08	080	130	071	166	110	078	137	078					
6 A	2	10101	08	024	034	025	010	021	014	051	018					
6 B	2	10101	03	037	031	007										
6 C	2	10101	12	041	059	051	031	035	049	033	033	052	030	034	052	
6 U	2	10101	10	026	020	048	013	042	082	037	024	042	011			

INVESTIGATOR: DUANE BENJIT (01), TEST: A (01)

DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS

SIX LEVELS: 1(CONTROL), 2(LOWEST), 6(HIGHEST); 4 REPS EA (A,B,C,D)

SEE DATA SHEET FOR NOTES

MEASURED CONCENTRATIONS OF TOXICANT (MICRO G/L)

1 A	3	10101	061179	0.06
1 A	3	20101	062579	0.08
1 A	3	30101	070979	0.15
1 B	3	10101	060679	0.00
1 B	3	20101	061879	0.08
1 B	3	30101	070279	0.10
1 C	3	10101	060879	0.07
1 C	3	20101	062179	0.06
1 C	3	30101	070579	0.13
1 D	3	10101	061479	0.06
1 D	3	20101	062879	0.10
2 A	3	10101	061179	1.68
2 A	3	20101	062579	1.59
2 A	3	30101	070979	2.00
2 B	3	10101	060679	1.83
2 B	3	20101	061479	1.69
2 B	3	30101	061879	1.27
2 B	3	40101	070279	1.58
2 C	3	10101	060879	1.70
2 C	3	20101	061879	1.40
2 C	3	30101	062179	1.52
2 C	3	40101	070579	2.15
2 D	3	10101	061479	1.84
2 D	3	20101	062879	1.84
3 A	3	10101	061179	3.03
3 A	3	20101	062579	2.69
3 A	3	30101	070979	4.00
3 B	3	10101	060679	3.26
3 B	3	20101	061479	3.58
3 B	3	30101	061879	2.81
3 B	3	40101	070279	3.32
3 C	3	10101	060879	2.85
3 C	3	20101	061879	2.36
3 C	3	30101	062179	2.72
3 C	3	40101	070579	3.61
3 D	3	10101	061479	3.79
3 D	3	20101	062879	3.44
4 A	3	10101	060879	5.99
4 A	3	20101	061179	6.37
4 A	3	30101	062579	5.57
4 A	3	40101	070979	7.30
4 B	3	10101	060679	7.06
4 B	3	20101	060679	6.83
4 B	3	30101	061179	7.19
4 B	3	40101	061879	5.91
4 B	3	50101	062579	6.09
4 B	3	60101	070279	6.49

4 B 3 70101	070579	7.92
4 B 3 80101	070979	7.80
4 C 3 10101	060879	5.66
4 C 3 20101	061179	6.21
4 C 3 30101	062179	5.23
4 C 3 40101	062179	5.14
4 C 3 50101	062579	5.44
4 C 3 60101	070279	5.95
4 C 3 70101	070579	7.02
4 C 3 80101	070979	7.50
4 U 3 10101	061179	6.72
4 D 3 20101	061479	7.57
4 U 3 30101	062579	5.93
4 D 3 40101	062879	7.76
4 D 3 50101	070979	8.40
5 A 3 10101	061179	13.40
5 A 3 20101	062579	12.00
5 A 3 30101	070979	18.00
5 B 3 10101	060679	17.20
5 B 3 20101	061879	10.80
5 B 3 30101	070279	11.80
5 C 3 10101	060879	10.60
5 C 3 20101	062179	10.70
5 C 3 30101	070579	13.10
5 D 3 10101	061479	13.60
5 D 3 20101	062879	15.30
6 A 3 10101	061179	24.10
6 A 3 20101	062579	23.30
6 A 3 30101	070979	30.00
6 B 3 10101	060679	29.80
6 B 3 20101	061879	28.80
6 B 3 30101	070279	32.70
6 C 3 10101	060879	24.20
6 C 3 20101	062179	24.20
6 C 3 30101	070579	33.40
6 D 3 10101	061479	21.20
6 D 3 20101	062879	23.30

INVESTIGATOR: DEFOE (02), TEST: C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS : 1(CONTROL), 2(LOWEST), ..., 6(HIGHEST); 2 REPS EA (A,B)
 # EMBRYOS TESTED, # LIVE AFTER HATCH, # NORMAL FRY AFTER HATCH
 # FRY TESTED, # ALIVE AT END, # NORMAL AT END

1 A 1 10201	50	29	29	20	20	20
1 B 1 10201	50	31	31	20	20	20
2 A 1 10201	50	18	17	20	20	20
2 B 1 10201	50	31	31	20	20	20
3 A 1 10201	50	30	30	20	20	20
3 B 1 10201	50	33	33	20	19	18
4 A 1 10201	50	34	34	21	21	21
4 B 1 10201	50	29	29	20	19	19
5 A 1 10201	50	28	28	20	16	16
5 B 1 10201	50	33	33	20	15	15
6 A 1 10201	50	31	30	20	00	00
6 B 1 10201	50	31	00	20	00	00

INVESTIGATOR: DEFOE (02), TEST: C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS : 1(CONTROL), 2(LOWEST), ..., 6(HIGHEST); 2 REPS EA (A,B)
 INDIVIDUAL WEIGHTS (MG) OF ALL FISH ALIVE AT END OF TEST (2 CARDS OR LESS/CELL)
 NUMBER OF WEIGHTS - LIST OF WEIGHTS

1 A 2 10201 20	159	139	105	170	166	179	124	132	068	095	113	180	092
1 A 2 20201 20	149	153	174	138	153	172	139						
1 B 2 10201 20	011	147	172	173	174	151	200	100	124	238	173	145	170
1 B 2 20201 20	180	155	150	089	230	135	071						
2 A 2 10201 20	124	143	012	109	141	142	101	124	162	129	153	187	128
2 A 2 20201 20	191	149	124	103	103	192	080						
2 B 2 10201 20	091	192	195	214	193	197	095	220	116	164	162	200	190
2 B 2 20201 20	161	206	200	190	192	137	181						
3 A 2 10201 20	171	148	095	179	096	160	129	157	179	153	102	146	137
3 A 2 20201 20	131	159	144	152	112	177	198						
3 B 2 10201 18	170	153	151	165	131	134	149	152	162	137	120	129	131
3 B 2 20201 18	100	096	129	155	047								
4 A 2 10201 21	097	133	089	122	189	162	089	135	165	144	122	152	073
4 A 2 20201 21	193	144	145	145	139	226	150	045					
4 B 2 10201 19	115	111	070	099	133	084	132	114	064	102	090	091	131
4 B 2 20201 19	177	053	124	090	109	119							
5 A 2 10201 16	037	056	072	047	056	053	037	050	044	007	024	047	049
5 A 2 20201 16	037	004	059										
5 B 2 10201 15	029	050	039	055	035	033	057	036	056	045	050	063	003
5 B 2 20201 15	045	038											
6 A 2 10201 00													
6 B 2 10201 00													

INVESTIGATOR: DEFOE (02), TEST: C
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS : 1(CONTROL), 2(LOWEST), ..., 5(HIGHEST); 2 REPS EA (A,B)
 MEASURED CONCENTRATIONS OF TOXICANT (NOTE NO UNITS GIVEN)

1 A 3 10201	090579	000.068
1 A 3 20201	090779	000.065
1 A 3 30201	091079	000.074
1 A 3 40201	091279	000.116
1 A 3 50201	091479	000.068
1 A 3 60201	091979	000.023
1 A 3 70201	092979	000.024
1 A 3 80201	100279	000.033
1 B 3 10201	090679	000.078
1 B 3 20201	091179	000.054

1 B 3 30201	091379	000.045
1 B 3 40201	091779	000.018
1 B 3 50201	092079	000.042
1 B 3 60201	092779	000.026
1 B 3 70201	100179	000.024
1 B 3 80201	100379	000.033
2 A 3 10201	090579	001.95
2 A 3 20201	090779	1.82
2 A 3 30201	091079	2.54
2 A 3 40201	091279	2.24
2 A 3 50201	091479	1.84
2 A 3 60201	091979	1.83
2 A 3 70201	092979	1.96
2 A 3 80201	100279	2.02
2 B 3 10201	090679	1.78
2 B 3 20201	091179	2.34
2 B 3 30201	091379	2.15
2 B 3 40201	091779	1.94
2 B 3 50201	092079	1.99
2 B 3 60201	092779	1.54
2 B 3 70201	100179	2.01
2 B 3 80201	100379	2.01
3 A 3 10201	090579	7.47
3 A 3 20201	090779	5.66
3 A 3 30201	091079	4.81
3 A 3 40201	091279	4.28
3 A 3 50201	091479	6.16
3 A 3 60201	091979	6.55
3 A 3 70201	092819	4.73
3 A 3 80201	100279	6.03
3 B 3 10201	090679	5.22
3 B 3 20201	091179	7.06
3 B 3 30201	091379	8.40
3 B 3 40201	091779	4.73
3 B 3 50201	092079	6.20
3 B 3 60201	092779	4.54
3 B 3 70201	100179	6.99
3 B 3 80201	100379	6.74
4 A 3 10201	090579	14.17
4 A 3 20201	090779	13.32
4 A 3 30201	091079	16.87
4 A 3 40201	091279	17.35
4 A 3 50201	091479	12.39
4 A 3 60201	091979	14.19
4 A 3 70201	092979	14.99
4 A 3 80201	100279	13.33
4 B 3 10201	090679	13.55
4 B 3 20201	091179	17.12
4 B 3 30201	091379	16.57
4 B 3 40201	091779	14.76
4 B 3 50201	092079	13.73
4 B 3 60201	092779	14.37
4 B 3 70201	100179	15.10
4 B 3 80201	100379	14.14
5 A 3 10201	090579	43.71
5 A 3 20201	090779	43.15
5 A 3 30201	091079	55.80
5 A 3 40201	091279	52.20
5 A 3 50201	091479	42.46
5 A 3 60201	091979	46.11
5 A 3 70201	092979	48.36
5 A 3 80201	100379	51.63
5 B 3 10201	090679	48.21

5 A 3 20201	021172	49.50
5 A 3 30201	021372	51.50
5 B 3 40201	091772	43.24
5 A 3 50201	092079	44.20
5 B 3 60201	022772	43.20
5 A 3 70201	100172	51.49
5 A 3 80201		51.50
6 A 3 10201	020572	135.4
6 A 3 20201	020772	117.47
6 A 3 30201	021072	154.25
6 A 3 40201	091279	156.75
6 A 3 50201	021472	131.0
6 B 3 10201	020572	144.75
6 B 3 20201	021172	162.75
6 B 3 30201	021372	153.0

INVESTIGATOR: HOLCOMBE & PIERCE (1961), TEST: D (193)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 4 REPS EA (A,B,C,D)
 # EMBRYOS TESTED, # LIVE AFTER HATCH, # NORMAL FRY AFTER HATCH

# FRY TESTED	# ALIVE AT END	# NORMAL AT END
1 A 1 10603	50	33
1 B 1 10603	50	29
1 C 1 10603	50	38
1 D 1 10603	50	30
2 A 1 10603	50	31
2 B 1 10603	50	36
2 C 1 10603	50	34
2 D 1 10603	50	29
3 A 1 10603	50	35
3 B 1 10603	50	38
3 C 1 10603	50	26
3 D 1 10603	50	38
4 A 1 10603	50	30
4 B 1 10603	50	29
4 C 1 10603	50	27
4 D 1 10603	50	36
5 A 1 10603	50	32
5 B 1 10603	50	31
5 C 1 10603	50	36
5 D 1 10603	50	34
6 A 1 10603	50	36
6 B 1 10603	50	34
6 C 1 10603	50	25
6 D 1 10603	50	29

INVESTIGATOR: HOLCOMBE & PHIPPS (1961). TEST: D (33)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1 (CONTROL), 2 (LOWEST), ..., 6 (HIGHEST); 4 REPS EA (A,B,C,D)
 INDIVIDUAL WEIGHTS (MG) OF ALL FISH ALIVE AT END OF TEST (2 CARDS OR LESS/CELL)
 NUMBER OF WEIGHTS & LIST OF WEIGHTS

1 A 2 10603 23	121	104	114	143	119	144	136	167	110	090	118	145	109
1 A 2 20603 23	140	129	167	137	156	132	128	147	120	125			
1 B 2 10603 23	147	173	147	177	117	115	149	177	136	162	138	185	121
1 B 2 20603 23	141	116	176	141	127	147	092	165	127	103			
1 C 2 10603 24	156	127	137	155	095	159	134	146	152	150	045	096	067
1 C 2 20603 24	173	190	129	159	098	104	109	116	134	136	095		
1 D 2 10603 24	129	175	177	137	161	120	142	108	134	099	119	144	130
1 D 2 20603 24	123	108	127	135	147	105	137	121	103	124	116		
2 A 2 10603 22	144	101	135	115	117	127	150	131	178	103	120	146	132
2 A 2 20603 22	115	112	176	109	137	137	110	116	116				
2 B 2 10603 24	129	160	154	131	147	094	119	165	145	147	168	144	124
2 B 2 20603 24	145	133	115	141	143	112	149	113	084	122	154		
2 C 2 10603 23	151	145	113	132	157	181	146	136	140	131	119	176	140
2 C 2 20603 23	155	163	149	117	150	123	123	137	103	147			
2 D 2 10603 23	177	211	125	147	134	124	153	165	241	110	140	151	128
2 D 2 20603 23	102	111	123	137	142	115	128	153	154	101			
3 A 2 10603 23	133	145	125	174	120	177	149	135	125	074	094	111	170
3 A 2 20603 23	153	166	107	137	144	119	145	117	137	015			
3 B 2 10603 25	131	120	193	127	145	137	123	154	102	131	010	124	156
3 B 2 20603 25	137	149	141	123	113	141	117	124	079	095	074	079	
3 C 2 10603 20	168	137	173	135	125	113	140	120	140	153	142	120	155
3 C 2 20603 20	143	133	140	122	122	113	099						
3 D 2 10603 24	117	137	144	114	137	167	151	107	139	143	124	107	114
3 D 2 20603 24	143	144	139	113	115	129	103	102	125	068	083		
4 A 2 10603 21	110	149	147	151	118	094	117	152	134	142	099	118	111
4 A 2 20603 21	124	102	175	095	098	042	012	020					
4 B 2 10603 20	186	130	133	150	141	049	076	122	195	149	100	118	140
4 B 2 20603 20	115	123	133	110	127	072	094						
4 C 2 10603 21	177	171	119	124	137	115	163	114	135	153	109	084	136
4 C 2 20603 21	134	110	101	094	104	090	076	049					
4 D 2 10603 25	130	136	125	142	120	122	105	104	139	115	129	123	114
4 D 2 20603 25	134	093	053	087	113	082	051	041	047	077	043	056	
5 A 2 10603 02	133	074											
5 A 2 20603 04	117	117	134	052									
5 C 2 10603 09	118	123	129	133	136	131	147	034	054				
5 D 2 10603 06	135	129	102	097	041	034							
6 A 2 10603 00													
6 B 2 10603 00													
6 C 2 10603 00													
6 D 2 10603 00													

INVESTIGATOR: HOLCOMBE & PHIPPS (1961), TEST: D (33)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1(CONTROL), 2(LWTEST), 6(HIGHEST); 4 REPS EA (A,B,C,D)
 MEASURED CONCENTRATIONS OF TOXICANT (MICRO G / L)

1 A 3 10603	0.		
1 B 3 10603	0.		
1 C 3 10603	0.		
1 D 3 10603	0.		
2 A 3 10603	022772	25.1	
2 A 3 20603	090479	26.7	
2 A 3 30603	091079	24.5	
2 A 3 40603	091779	28.1	
2 A 3 50603	092479	26.7	
2 B 3 10603	082372	26.0	
2 B 3 20603	090579	27.0	
2 B 3 30603	091172	26.1	
2 B 3 40603	091972	27.2	
2 B 3 50603	092572	27.0	
2 C 3 10603	082479	26.6	
2 C 3 20603	090579	27.9	
2 C 3 30603	091272	25.9	
2 C 3 40603	091972	27.7	
2 C 3 50603	092579	28.0	
2 D 3 10603	083072	27.9	
2 D 3 20603	090772	25.6	
2 D 3 30603	091372	25.5	
2 D 3 40603	092072	27.5	
2 D 3 50603	092772	25.7	
3 A 3 10603	082772	44.1	
3 A 3 20603	090479	40.1	
3 A 3 30603	091072	41.2	
3 A 3 40603	091772	47.3	
3 A 3 50603	092472	44.7	
3 B 3 10603	082772	43.3	
3 B 3 20603	091579	47.2	
3 B 3 30603	092172	45.5	
3 B 3 40603	091972	46.3	
3 B 3 50603	092572	45.6	
3 C 3 10603	082972	30.9	
3 C 3 20603	090579	45.1	
3 C 3 30603	091279	40.1	
3 C 3 40603	091979	40.7	
3 C 3 50603	092572	47.5	
3 D 3 10603	083072	45.7	
3 D 3 20603	090772	47.1	
3 D 3 30603	091372	46.7	
3 D 3 40603	092072	42.4	
3 D 3 50603	092772	40.0	
4 A 3 10603	082772	49.7	
4 A 3 20603	090479	75.5	
4 A 3 30603	091079	48.2	
4 A 3 40603	091779	75.4	
4 A 3 50603	092472	73.7	
4 B 3 10603	082972	43.1	
4 B 3 20603	090579	72.6	
4 B 3 30603	091172	71.2	
4 B 3 40603	091972	73.2	
4 B 3 50603	092572	60.2	
4 C 3 10603	082972	45.0	
4 C 3 20603	090579	70.0	
4 C 3 30603	091272	74.7	
4 C 3 40603	091972	72.2	
4 C 3 50603	092572	73.1	
4 D 3 10603	083072	75.0	
4 D 3 20603	090772	75.5	
4 D 3 30603	091372	75.2	
4 D 3 40603	092072	72.2	
4 D 3 50603	092772	45.4	
5 A 3 10603	082772	113.	
5 A 3 20603	090479	143.	
5 A 3 30603	091072	141.	
5 A 3 40603	091779	132.	
5 A 3 50603	092472	131.	
5 B 3 10603	092472	136.	
5 B 3 20603	090579	130.	
5 B 3 30603	091172	131.	
5 B 3 40603	091372	127.	
5 B 3 50603	092572	120.	
5 C 3 10603	082772	110.	
5 C 3 20603	090579	112.	
5 C 3 30603	091272	149.	
5 C 3 40603	091972	122.	
5 C 3 50603	092572	117.	
5 D 3 10603	083072	123.	
5 D 3 20603	090772	125.	
5 D 3 30603	091372	130.	
5 D 3 40603	092072	115.	
5 D 3 50603	092772	124.	
6 A 3 10603	082772	233.	
6 A 3 20603	090479	236.	
6 A 3 30603	091079	221.	
6 A 3 40603	091772	213.	
6 A 3 50603	092472	210.	
6 B 3 10603	082772	230.	
6 B 3 20603	090579	243.	
6 B 3 30603	091172	212.	
6 B 3 40603	091872	224.	
6 B 3 50603	092579	135.	
6 C 3 10603	082772	230.	
6 C 3 20603	090479	216.	
6 C 3 30603	091272	244.	
6 C 3 40603	091972	210.	
6 C 3 50603	092572	206.	
6 D 3 10603	083072	235.	
6 D 3 20603	090772	222.	
6 D 3 30603	091372	233.	
6 D 3 40603	092072	234.	
6 D 3 50603	092772	221.	

APPENDIX AVII.* THEORETICAL BASES OF CHI SQUARE AND PROBIT BASED TESTS
OF TANK TO TANK VARIATION WITHIN TREATMENT GROUPS

A. Chi Square Test for Heterogeneity of Tanks Within Treatment Groups

Suppose that there are I treatment (or control) groups and J tanks per treatment group. Let p_i denote the probability of death within the i-th group, $i = 1, 2, \dots, I$ assuming homogeneity of tanks within groups. The heterogeneity chi square procedure tests the hypothesis

$$H_0: p_1 = p_2 = \dots = p_I = p.$$

Let N_{ij} , X_{ij} denote the number of fish and number of dead fish respectively in the j-th tank of the i-th group. Let $X_{i+} = \sum_j X_{ij}$, $X_{++} = \sum_i \sum_j X_{ij}$, $N_{i+} = \sum_j N_{ij}$, $N_{++} = \sum_i \sum_j N_{ij}$, $\hat{p}_{ij} = X_{ij}/N_{ij}$, $\hat{p}_{i+} = X_{i+}/N_{i+}$, $\hat{p} = X_{++}/N_{++}$.

The chi test of homogeneity based on each tank separately is based on the statistic

$$\chi_I^2 \equiv \sum_{i=1}^I \sum_{j=1}^J \frac{(X_{ij} - N_{ij}\hat{p})^2}{N_{ij}\hat{p}(1-\hat{p})} \quad \text{with } IJ - 1 \text{ d.f.}$$

The chi square test of homogeneity based on tanks pooled within treatment groups is based on the statistic

$$\chi_{II}^2 \equiv \sum_{i=1}^I \frac{(X_{i+} - N_{i+}\hat{p})^2}{N_{i+}\hat{p}(1-\hat{p})} \quad \text{with } I - 1 \text{ d.f.}$$

The difference of these two statistics, $\chi_I^2 - \chi_{II}^2$, is

$$\sum_{i=1}^I \sum_{j=1}^J \frac{(X_{ij} - N_{ij}\hat{p}_{i+})^2}{N_{ij}\hat{p}_{i+}(1-\hat{p}_{i+})} \quad \text{with } I(J - 1) \text{ d.f.}$$

This difference is thus a test of homogeneity of response rates across tanks within treatment groups and has a nominal chi square distribution with $I(J - 1)$ degrees of freedom if such homogeneity exists. Note however that the weights in the denominator of this "chi square" statistic have been calculated under the null hypothesis that $p_1 = p_2 = \dots = p_I = p$. Thus if the response probabilities differ among treatment groups, as we have seen in the case with respect to fry mortality, then the weights are incorrect and the nominal null distribution is inappropriate. Thus this procedure leaves something to be desired.

*Appendix AVII is the appendix for Section VII.

B. Probit Model Based Test for Heterogeneity of Tanks Within Treatment Groups

Let X_{ij} , N_{ij} , X_{i+} , N_{i+} , \hat{p}_{ij} , \hat{p}_i have the same interpretation as in subsection A. Let $\hat{\phi}_i \equiv \Phi(\alpha + \beta C_i)$ denote the response probability in the i -th group where C_i is some function of the toxicant concentration in the i -th group, Φ is the standard normal c.d.f. and α , β are model parameters to be estimated from the data. Let $\hat{\phi}_i \equiv \Phi(\hat{\alpha} + \hat{\beta} C_i)$ denote the maximum likelihood estimate of ϕ_i .

Finney's suggestion is to compare test statistics for lack of fit to the probit model based on each tank separately and based on tanks pooled within treatment groups. The lack of fit test based on each tank separately is

$$\chi^2_I \equiv \sum_{i=1}^I \sum_{j=1}^J \frac{(X_{ij} - N_{ij} \hat{\phi}_i)^2}{N_{ij} \hat{\phi}_i (1 - \hat{\phi}_i)} \quad \text{with } IJ - 2 \text{ d.f.}$$

The lack of fit test based on tanks pooled within treatment groups is

$$\chi^2_{II} \equiv \sum_{i=1}^I \frac{(X_{i+} - N_{i+} \hat{\phi}_i)^2}{N_{i+} \hat{\phi}_i (1 - \hat{\phi}_i)} \quad \text{with } I - 2 \text{ d.f.}$$

The difference is these two statistics,

$$\chi^2_I - \chi^2_{II} = \sum_{i=1}^I \sum_{j=1}^J \frac{(X_{ij} - N_{ij} \hat{\phi}_i)^2}{N_{ij} \hat{\phi}_i (1 - \hat{\phi}_i)} \quad \text{with } I(J - 1) \text{ d.f.}$$

This difference is again a test of homogeneity of response rates across tanks within treatment groups and has a nominal chi square distribution with $I(J - 1)$ degrees of freedom if such homogeneity exists. Note that the weights in the denominator of the statistic are based on the fitted probit model and depend for their validity on the validity of this model.

APPENDIX AVIII.1* OUTPUT FROM EXAX2 COMPUTER PROGRAM

* Appendix AVIII.1 is an appendix for Section VII.1.

***** DEFECT COMPOUND C DEFECT MORTALITY *****

CASE 1

CONFIDENCY TABLE

COL 1	ALIVE	DEAD	TOTAL
COL 2	28.000	22.000	50.000
	31.000	19.000	50.000
TOTAL	ALIVE = 59.000	TOTAL DEAD = 41.000	GRAND TOTAL = 100.000

XSCORES = 0.37205 CUTOFF = 5.00000

COL 1	ALIVE	DEAD	TOTAL
COL 2	29.500	20.500	50.000
	29.500	20.500	50.000
TOTAL	ALIVE = 59.000	TOTAL DEAD = 41.000	GRAND TOTAL = 100.000

ASYMPTOTIC CHI SQUARED USED
OBSERVED CHI SQUARED = 0.37205
OBSERVED SIGNIFICANCE LEVEL = 0.54189
VY = -2LN(A1) = 1.22540
VY = 1(-2LN(A1)) = 2.00000
VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.1 EXAX2 output

 DEFEC COMPOUND C EMERYD MORTALITY

CASE 2

CONTINGENCY TABLE

COL	1	ALIVE	DEAD	TOTAL			
COL	2	19.000	32.000	50.000			
TOTAL		31.000	19.000	50.000			
ALIVE	=	49.000	TOTAL	DEAD	=	51.000	GRAND TOTAL = 100.000

EXPECTED FREQUENCIES							
COL	1	ALIVE	DEAD	TOTAL			
COL	2	24.500	25.500	50.000			
COL	3	24.500	25.500	50.000			
ALIVE	=	49.000	TOTAL	DEAD	=	51.000	GRAND TOTAL = 100.000

CUTOFF = 6.76271							
XSQRS = 6.00000							

EXPECTED FREQUENCIES

XSQURS = 6.76271 CUTOFF = 6.00000

ASYMPTOTIC CHI SQUARED USED
 OBSERVED CHI SQUARED = 6.76271
 OBSERVED SIGNIFICANCE LEVEL = .000931
 YV = -2LN(A1) = 9.35371
 EV = 2(-2LN(A1)) = 2.00000
 VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.2 EXAX2 output

DEFOE COMPOUND C EMBAYO MORTALITY

CASE 3

CONTINGENCY TABLE

	ALIVE	DEAD	TOTAL
COL 1	30.000	20.000	50.000
COL 2	33.000	17.000	50.000
TOTAL	63.000	37.000	100.000

TOTAL ALIVE = 63.000 , TOTAL DEAD = 37.000 , GRAND TOTAL = 100.000

EXPECTED FREQUENCIES

XSQARS = 0.38610 CUTOFF = 9.000000

	ALIVE	DEAD	TOTAL
COL 1	31.500	18.500	50.000
COL 2	31.500	18.500	50.000

TOTAL ALIVE = 63.000 , TOTAL DEAD = 37.000 , GRAND TOTAL = 100.000

ASYMPTOTIC CHI SQUARED USED

RESERVED CHI SQUARED = 0.38610

RESERVED SIGNIFICANCE LEVEL = .053436

YY = 2LN(AI) = 1.25333

YY = 2LN(AI) = 2.00000

VARY = VAR(-2LN(AI)) = 4.00000

Figure VIII.3 EXAX2 output

DEFOE COMPOND C EMERYE MORTALITY

CASE 4

CONTINGENCY TABLE

		ALIVE		DEAD	TOTAL	
COL	ROW	1	2		1	2
1	1	34.000		16.000	50.000	
2	2	29.000		21.000	50.000	
TOTAL		63.000		37.000	100.000	

		ALIVE		DEAD	TOTAL	
COL	ROW	1	2		1	2
1	1	31.500		18.500	50.000	
2	2	31.500		19.500	50.000	
TOTAL		63.000		37.000	100.000	

		EXPECTED FREQUENCIES		XSQURS =	CUTOFF =
				1.07250	5.00.000

XSQURS = 1.07250 CUTOFF = 5.00000

EXPECTED FREQUENCIES

		ALIVE		DEAD	TOTAL	
COL	ROW	1	2		1	2
1	1	31.500		18.500	50.000	
2	2	31.500		19.500	50.000	
TOTAL		63.000		37.000	100.000	

ASYMPTOTIC CHI SQUARED USED

OBSERVED - D CHI SQUARED = 1.07250
 OBSERVED SIGNIFICANCE LEVEL = .A1 = 0.30039
 YY = -2LN(A1) = 2.40841
 Y = 5(-2LN(A1)) = 2.00000
 VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.4 EXAX2 output

DEFOE Compound C EMBRYO MORTALITY

CASE 5

CONTINGENCY TABLE

COL	1	2	ALIVE	DEAD	TOTAL
COL	1	2	28.000	22.000	50.000
COL	2	1	33.000	17.000	50.000
TOTAL	ALIVE	=	61.000	TOTAL	DEAD = 39.000 , GRAND TOTAL = 100.000

COL	1	2	ALIVE	DEAD	TOTAL
COL	1	2	20.500	19.500	40.000
COL	2	1	30.500	19.500	50.000
TOTAL	ALIVE	=	61.000	TOTAL	DEAD = 39.000 , GRAND TOTAL = 100.000

EXPECTED FREQUENCIES

COL	1	2	ALIVE	DEAD	TOTAL
COL	1	2	20.500	19.500	40.000
COL	2	1	30.500	19.500	50.000
TOTAL	ALIVE	=	61.000	TOTAL	DEAD = 39.000 , GRAND TOTAL = 100.000

ASYMPTOTIC CHI SQUARED USED

OBSERVED CHI SQUARED = 1.05086
 OBSERVED SIGNIFICANCE LEVEL = 0.30531
 P = 2LN(A1) = 2.37285
 P = 2(-2LN(A1)) = 2.00000
 VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.5 EXAX2 output

DETOX COMPOUND C EMBRYO MORTALITY

CASE 6

CONTINGENCY TABLE

COL	1	2	ALIVE	DEAD	TOTAL
COL	1	2	31.000	19.000	50.000
COL	2	2	31.000	19.000	50.000
TOTAL	ALIVE =	62.000	TOTAL	DEAD =	38.000
				GRAND TOTAL =	100.000
EXPECTED FREQUENCIES					
			XSDURS =	0.0	CUTOFF = 5.00000
COL	1	2	ALIVE	DEAD	TOTAL
COL	1	2	21.000	19.000	50.000
COL	2	2	31.000	19.000	50.000
TOTAL	ALIVE =	62.000	TOTAL	DEAD =	38.000
				GRAND TOTAL =	100.000

ASYMPTOTIC CHI SQUARED USED
 OBSERVED CHI SQUARED = 0.0
 OBSERVED SIGNIFICANCE LEVEL = AI = 1.00000
 YY = -2LN(AI) = 0.0
 FY = (-2LN(AI)) = 2.00000
 VARY = VAR(-2LN(AI)) = 4.00000

6 INDEPENDENT TESTS COMBINED

SY = SUM(VY) = 16.61076
 YMU = SUM(VY) = 12.00000
 SVARY = SUM(VARY) = 24.00000

Z = (SY - SORT(YMU))/SQRT(SVARY/(4*YMU)) = 0.96493
 UNDER H0 Z HAS A STANDARD NORMAL DISTRIBUTION

 DEFEC COMPOND C MY MORTALITY

CASE 1

CONTINGENCY TABLE

COL	1	2	TOTAL	ALIVE	DEAD	TOTAL
COL	1	2	TOTAL	ALIVE	DEAD	TOTAL
	20.000	0.0	20.000	20.000	0.0	20.000
	20.000	0.0	20.000	20.000	0.0	20.000
TOTAL	ALIVE = 40.000	TOTAL DEAD = 0.0	GRAND TOTAL = 40.000			

TABLE IS DEGENERATE

OBSERVED SIGNIFICANCE LEVEL = AI = 1.0

YY = -2LN(AI) = 0.0

EY = E(-2LN(AI)) = 0.0

VARY = VAR(-2LN(AI)) = 0.0

Figure VIII.7 EXAX2 output

 DEFE COMPOUND FRY MORTALITY

CASE 2

CONTINGENCY TABLE

	ALIVE	DEAD	TOTAL
COL 1	20.000	0.0	20.000
COL 2	20.000	0.0	20.000
TOTAL	40.000	0.0	40.000

TABLE IS DEGENERATE

OBSERVED SIGNIFICANCE LEVEL = AI = 1.0

YY = $-2\text{LN}(AI)$ = 0.0

EV = $E(-2\text{LN}(AI))$ = 0.0

VARY = $\text{VAR}(-2\text{LN}(AI))$ = 0.0

Figure VIII.8 EXAX2 output

 DEFOE COMPOUND C FRY MORTALITY

CASE 3

CONTINGENCY TABLE

COL	1	2	ALIVE	DEAD	TOTAL					
COL	1	2	20.000	0.0	20.000					
COL	2	1	19.000	2.000	21.000					
TOTAL	ALIVE	=	38.000	TOTAL	DEAD	=	2.000	GRAND TOTAL	=	40.000

EXPECTED FREQUENCIES

COL	1	2	ALIVE	DEAD	TOTAL					
COL	1	2	19.000	1.000	20.000					
COL	2	1	19.000	1.000	20.000					
TOTAL	ALIVE	=	38.000	TOTAL	DEAD	=	2.000	GRAND TOTAL	=	40.000

EXACT X² DISTRIBUTION
 X² = 0.0
 P-VALUE = 0.51242
 SIG = 1.00000
 2.10526 0.48718 0.48718

EXACT CHI-SQUARED DISTRIBUTION USED
 OBSERVED CHI-SQUARED = 2.10526
 OBSERVED SIGNIFICANCE LEVEL = 0.48718
 YY = -2LN(A1) = 1.43825
 XY = 2(-2LN(A1)) = 0.70060
 VARY = VAR(-2LN(A1)) = 0.51680

Figure VIII.9 EXAX2 output

***** OESOE COMPANYNO C PRY MORTALITY *****

CASE 4

CONTINGENCY TABLE

COL 1	ALIVE	DEAD	TOTAL
COL 2	21.000	0.0	21.000
	19.000	1.000	20.000
TOTAL	ALIVE = 40.000	TOTAL DEAD = 1.000	GRAND TOTAL = 41.000
EXPECTED FREQUENCIES			
	XSQURS = 1.07625		CUTOFF = 5.00000

EXACT XSO DISTRIBUTION
 XSO: PR08 SIG
 0.27619 0.51220 1.00000
 1.07625 0.48780 0.49780

EXACT CHI SQUARED DISTRIBUTION USED
 OBSERVED CHI-SQUARED = 1.07625
 OBSERVED SIGNIFICANCE LEVEL = 0.49780
 YY = 1.43568
 -Y = 1.07625
 VARY = VAR(-2LN(AI)) = 0.70023
 -Y = 0.51489

DEFOE COMPOUND C FRY MORTALITY

CASE 8

CONFIDENCY TABLE

CUL	1	ALIVE	DEAD	TOTAL
CUL	2	16.000	4.000	20.000
		15.000	5.000	20.000
TOTAL	ALIVE =	31.000	TOTAL DEAD =	9.000
			GRAND TOTAL =	40.000

XSQBS = 0.14337

CUTOFF = 5.000000

EXPECTED FREQUENCIES

CUL	1	ALIVE	DEAD	TOTAL
CUL	2	15.500	4.500	20.000
		15.500	4.500	20.000
TOTAL	ALIVE =	31.000	TOTAL DEAD =	9.000
			GRAND TOTAL =	40.000

EXACT XSO DISTRIBUTION

XSO	PROB	SIG
0.14337	0.54442	1.00000
1.29032	0.32319	0.45058
3.53423	0.10773	0.13739
7.02500	0.01843	0.01660
11.61290	0.00123	0.00123

XACT CHI SQUARED DISTRIBUTION USED

OR SERV TO CHI SQUARED = 0.14337
 OR SERV TO SIGNIFICANT LEVEL = 0.00000
 VV = -2LN(AI) = -0.00000
 FY = (-2LN(AI)) = 1.17056
 VARY = VAR(-2LN(AT)) = 0.75430

Figure VIII.11 EXAX2 output

DEFOE COMPBAND G FRY MORTALITY

CASE 6

CONTINGENCY TABLE

COL	AL	ALIVE	DEAD	TOTAL
1	0.0	20.000	20.000	40.000
2	0.0	20.000	20.000	40.000
TOTAL	ALIVE = 0.0	TOTAL DEAD = 40.000	GRAND TOTAL = 40.000	

TABLE IS DEGENERATE

OBSERVED SIGNIFICANCE LEVEL = AI = 1.0

YY = -2LN(AI) = 0.0

EY = 1 - 2LN(AI) = 0.0

VARY = VARI(-2LN(AI)) = 0.0

6 INDEPENDENT TESTS COMBINED

SY = SUM(YI) = 2.87200

YMU = SUM(EI) = 2.52157

SVARY = SUM(VARI) = 3.78617

Z = (SUM(SY) - SUM(YMU)) / SQRT(SVARY / (4 * YMU)) = 0.17516

UNDER H0 Z HAS STANDARD NORMAL DISTRIBUTION

Figure VIII.12 EXAX2 output

HOLCOMBE AND PHIPPS COMPOUND D ENERGY MORTALITY

GROUP 1

CONTINGENCY TABLE

	1	2	3	4	TOTAL
DEAD	14.000	14.000	14.000	14.000	56.000
ALIVE	36.000	36.000	36.000	36.000	144.000
TOTAL	50.000	50.000	50.000	50.000	200.000

EXPECTED FREQUENCIES

	1	2	3	4	TOTAL
DEAD	14.000	14.000	14.000	14.000	56.000
ALIVE	36.000	36.000	36.000	36.000	144.000
TOTAL	50.000	50.000	50.000	50.000	200.000

ASYMPTOTIC CHI SQUARED USED

UNEXPECTED CHAT SQUARE = 4.30769
 OBSERVED SIGNIF LEVEL = A = 0.2000
 Y = -2LN(A) = 2.99574
 Y = -2LN(A) = 2.99574
 VARY = VAR(-2LN(A)) = 4.00000

Figure VIII.13 EXAX2 output

RECOMBINE AND PHENOLS COMPOUND D EMBRYO MORTALITY

GROUP 2

CONTINGENCY TABLE

	ALIVE	DEAD	TOTAL
TRT 1	21.000	15.000	36.000
TRT 2	25.000	14.000	39.000
TRT 3	24.000	16.000	40.000
TRT 4	20.000	21.000	41.000
TOTAL	130.000	70.000	200.000

EXPECTED FREQUENCIES

	ALIVE	DEAD	TOTAL
TRT 1	32.500	17.500	50.000
TRT 2	32.500	17.500	50.000
TRT 3	32.500	17.500	50.000
TRT 4	32.500	17.500	50.000
TOTAL	130.000	70.000	200.000

ASYMPTOTIC CHI-SQUARED USED:

OBSERVED CHI-SQUARED = 2.54945
 OBSERVED SIGNIFICANCE LEVEL = .11 = .946042
 P-VALUE = 1.52534
 P-VALUE = 1.52534
 P-VALUE = 1.52534
 P-VALUE = 1.52534

Figure VIII.14 EXAX2 output

HOLCOMBE AND PHIPPS COMPOUND EMBRYO MORTALITY

GROUP 3

CONTINGENCY TABLE

	1	2	TOTAL
IRI	1	1	2
IRI	2	1	2
IRI	2	1	2

TOTAL ALIVE = 137.000 , TOTAL DEAD = 63.000 , GRAND TOTAL = 200.000

EXPECTED FREQUENCIES

GUIDEF = 5.00000

	1	2	TOTAL
IRI	1	1	2
IRI	2	1	2
IRI	2	1	2

TOTAL ALIVE = 137.000 , TOTAL DEAD = 63.000 , GRAND TOTAL = 200.000

ASYMPTOTIC CHI SQUARED USED.

RESERVED CHI SQUARED = 9.96767

RESERVED SIGNIFICANCE LEVEL = 0.02972

YY = 7.00000

VARY = VARIANCE(1) = 4.00000

HOLCOMBE AND PHIPPS COMPOUND O ENEMY MORTALITY

GROUP 4

CONTINGENCY TABLE

	IRI	1	2	3	4	TOTAL	ALIVE	DEAD	TOTAL	DEAD	TOTAL	GRAND TOTAL
IRI	1	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	2	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	3	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	4	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
TOTAL	TOTAL	122.000	122.000	122.000	122.000	488.000	488.000	56.000	312.000	312.000	800.000	800.000

EXPECTED FREQUENCIES

	IRI	1	2	3	4	TOTAL	ALIVE	DEAD	TOTAL	DEAD	TOTAL	GRAND TOTAL
IRI	1	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	2	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	3	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
IRI	4	29.000	29.000	27.000	36.000	122.000	122.000	14.000	78.000	78.000	200.000	200.000
TOTAL	TOTAL	122.000	122.000	122.000	122.000	488.000	488.000	56.000	312.000	312.000	800.000	800.000

ASYMPTOTIC CHI SQUARED USED:
 OBSERVED CHI SQUARED = 3.73310
 OBSERVED SIGNIFICANCE LEVEL = 0.28936
 P = 0.28936
 SY = 0.28936
 VARY = VARI(2,1,1) = 4.00000

Figure VIII.16 EXAX2 output

HOLCOMBE AND PHIPS COMPOUND Q ENERGY MORTALITY

GROUP 5

CONTINGENCY TABLE

	1	2	3	4	TOTAL
TRI	21.000	18.000	19.000	14.000	50.000
TRI	21.000	18.000	19.000	14.000	50.000
TRI	21.000	18.000	19.000	14.000	50.000
TOTAL	67.000	67.000	67.000	67.000	206.000

EXPECTED FREQUENCIES

	1	2	3	4	TOTAL
TRI	33.250	16.750	16.750	16.750	50.000
TRI	33.250	16.750	16.750	16.750	50.000
TRI	33.250	16.750	16.750	16.750	50.000
TOTAL	123.000	67.000	67.000	67.000	304.000

ASYMPTOTIC CHI SQUARED USED.

UNEXPECTED CHI SQUARED = 1.32421
 UNEXPECTED CHI SQUARED. LEVEL = .0.72339
 YV = -2LN(4) = 0.62700
 YV = -2LN(4) = 0.62700
 VAR = VAR(2LN(4)) = 4.00000

Figure VIII.17 EXAX2 output

HOLCOMBE AND PHIPPS COMPANY O EMERGENCY MORTALITY

GROUP 6

CONTINGENCY TABLE

	1	2	3	4	TOTAL
1	24.000	14.000	14.000	31.000	83.000
2	24.000	14.000	14.000	31.000	83.000
3	24.000	14.000	14.000	31.000	83.000
4	24.000	14.000	14.000	31.000	83.000
TOTAL	96.000	56.000	56.000	126.000	334.000

TOTAL ALIVE = 96.000 , TOTAL DEAD = 126.000 , GRAND TOTAL = 334.000

EXPECTED FREQUENCIES

CUTOFF = 5.00001

	1	2	3	4	TOTAL
1	21.000	14.000	14.000	31.000	80.000
2	21.000	14.000	14.000	31.000	80.000
3	21.000	14.000	14.000	31.000	80.000
4	21.000	14.000	14.000	31.000	80.000
TOTAL	84.000	56.000	56.000	126.000	282.000

TOTAL ALIVE = 84.000 , TOTAL DEAD = 126.000 , GRAND TOTAL = 310.000

ASYMPTOTIC CHI SQUARED USED.

OBSERVED CHI SQUARED = 6.22185

DEGREES OF FREEDOM = 3 , P = 0.09869

Y = -2LN(L) = 4.63174

Y = -2LN(L) = 4.63174

WAS = 1 - LN(L) = 4.00000

6 INDEPENDENT TESTS COMBINED

SY = SUM(Y) = 9.77942

YMU = SUM(Y) = 12.00000

WAS = SUM(WAS) = 4.00000

Z = (SUM(Y) - SUM(YMU)) / SQRT(SUM(WAS) * SUM(WAS)) = 1.31060

UAD = 2 * HAD = 1.00000

WELCH AND SHARP COMPARE O RY MORTALITY

GROUP 1 CONTINGENCY TABLE

	ALIVE	DEAD	TOTAL
TKI 1	23.000	2.000	25.000
TKI 2	23.000	2.000	25.000
TKI 3	24.000	1.000	25.000
TKI 4	24.000	1.000	25.000
TOTAL	ALIVE = 94.000	TOTAL DEAD = 6.000	GRAND TOTAL = 100.000

CUTOFF = 5.00000

EXPECTED FREQUENCIES

	ALIVE	DEAD	TOTAL
TKI 1	23.500	1.500	25.000
TKI 2	23.500	1.500	25.000
TKI 3	23.500	1.500	25.000
TKI 4	23.500	1.500	25.000
TOTAL	ALIVE = 94.000	TOTAL DEAD = 6.000	GRAND TOTAL = 100.000

EXACT KSU DISTRIBUTION

KSU	PKUB	SIG
0.70922	0.28313	1.00000
2.12706	0.24119	0.11687
3.25610	0.34730	0.20568
4.38518	0.10622	0.12438
5.51422	0.03820	0.05217
6.64326	0.01331	0.01337
7.77230	0.00059	0.00059

EXACT DISTRIBUTION OF KSU STATISTIC USED

UNOBSERVED KSU = 0.70922
 OBSERVED SIGNIFICANCE LEVEL = A1 = 1.0000
 P1 = 0.00000
 P2 = 0.00000
 P3 = 0.00000
 P4 = 0.00000
 P5 = 0.00000
 P6 = 0.00000
 P7 = 0.00000
 P8 = 0.00000
 P9 = 0.00000
 P10 = 0.00000
 P11 = 0.00000
 P12 = 0.00000
 P13 = 0.00000
 P14 = 0.00000
 P15 = 0.00000
 P16 = 0.00000
 P17 = 0.00000
 P18 = 0.00000
 P19 = 0.00000
 P20 = 0.00000
 P21 = 0.00000
 P22 = 0.00000
 P23 = 0.00000
 P24 = 0.00000
 P25 = 0.00000
 P26 = 0.00000
 P27 = 0.00000
 P28 = 0.00000
 P29 = 0.00000
 P30 = 0.00000
 P31 = 0.00000
 P32 = 0.00000
 P33 = 0.00000
 P34 = 0.00000
 P35 = 0.00000
 P36 = 0.00000
 P37 = 0.00000
 P38 = 0.00000
 P39 = 0.00000
 P40 = 0.00000
 P41 = 0.00000
 P42 = 0.00000
 P43 = 0.00000
 P44 = 0.00000
 P45 = 0.00000
 P46 = 0.00000
 P47 = 0.00000
 P48 = 0.00000
 P49 = 0.00000
 P50 = 0.00000
 P51 = 0.00000
 P52 = 0.00000
 P53 = 0.00000
 P54 = 0.00000
 P55 = 0.00000
 P56 = 0.00000
 P57 = 0.00000
 P58 = 0.00000
 P59 = 0.00000
 P60 = 0.00000
 P61 = 0.00000
 P62 = 0.00000
 P63 = 0.00000
 P64 = 0.00000
 P65 = 0.00000
 P66 = 0.00000
 P67 = 0.00000
 P68 = 0.00000
 P69 = 0.00000
 P70 = 0.00000
 P71 = 0.00000
 P72 = 0.00000
 P73 = 0.00000
 P74 = 0.00000
 P75 = 0.00000
 P76 = 0.00000
 P77 = 0.00000
 P78 = 0.00000
 P79 = 0.00000
 P80 = 0.00000
 P81 = 0.00000
 P82 = 0.00000
 P83 = 0.00000
 P84 = 0.00000
 P85 = 0.00000
 P86 = 0.00000
 P87 = 0.00000
 P88 = 0.00000
 P89 = 0.00000
 P90 = 0.00000
 P91 = 0.00000
 P92 = 0.00000
 P93 = 0.00000
 P94 = 0.00000
 P95 = 0.00000
 P96 = 0.00000
 P97 = 0.00000
 P98 = 0.00000
 P99 = 0.00000
 P100 = 0.00000

Figure VIII.19 EXAX2 output

 HALLSOME AND PHIPPS COMPARED FRY MORTALITY

GROUP 3		CONTINGENCY TABLE			
		ALIVE	DEAD	TOTAL	
TKI	1	23.000	2.000	25.000	
TKI	2	23.000	0.000	23.000	
TKI	3	20.000	5.000	25.000	
TKI	4	24.000	1.000	25.000	
TOTAL	ALIVE =	92.000	TOTAL DEAD =	8.000	GRAND TOTAL = 100.000
EXCLUDED FREQUENCIES					
		ALIVE	DEAD	TOTAL	
TKI	1	23.000	2.000	25.000	
TKI	2	23.000	2.000	25.000	
TKI	3	23.000	2.000	25.000	
TKI	4	23.000	2.000	25.000	
TOTAL	ALIVE =	92.000	TOTAL DEAD =	8.000	GRAND TOTAL = 100.000
LUSOFF = 2.00000					

EXACT ASQ DISTRIBUTION

ASQ	PKUB	\$16
0.00	0.04254	1.00000
1.00090	0.33371	0.95097
2.00131	0.10660	0.61216
3.00081	0.23249	0.51626
4.00478	0.07342	0.26087
5.00347	0.09581	0.18745
6.00217	0.01784	0.09364
7.00370	0.05134	0.07574
8.00365	0.00516	0.02440
9.00201	0.00768	0.01424
10.00202	0.00714	0.01136
11.00434	0.00343	0.00422
12.00434	0.00377	0.00480
13.00434	0.00002	0.00002

EXACT DISTRIBUTION OF ASQ STATISTIC USED.
 OBSERVED ASQ = 10.0870
 OBSERVED SIGNIFICANCE LEVEL = A1 = 0.01514
 P1 = 2LN(A1) = 3.15948
 P2 = 2LN(A1) = 1.54297
 VARI = VARI(2LN(A1)) = 3.44140

Figure VIII.21 EXAX2 output

 NOLCOMBE AND PHIPS COMPOUND D FRY MORTALITY

GROUP		CONTINGENCY TABLE					
		ALIVE	DEAD	TOTAL			
TRI	1	21.000	4.000	25.000			
TRI	2	20.000	5.000	25.000			
TRI	3	21.000	4.000	25.000			
TRI	4	25.000	0.0	25.000			
TOTAL	ALIVE	= 87.000	TOTAL DEAD = 13.000	GRAND TOTAL = 100.000			

EXPECTED FREQUENCIES

CUTOFF = 5.00000

		ALIVE	DEAD	TOTAL			
TRI	1	21.750	3.250	25.000			
TRI	2	21.750	3.250	25.000			
TRI	3	21.750	3.250	25.000			
TRI	4	21.750	3.250	25.000			
TOTAL	ALIVE	= 87.000	TOTAL DEAD = 13.000	GRAND TOTAL = 100.000			

EXACT ASU DISTRIBUTION

ASU	PROB	SIG
0.26523	0.08658	1.00000
0.91259	0.18634	0.91342
1.61793	0.14230	0.72708
2.36721	0.13055	0.56478
3.07401	0.13044	0.42423
3.80195	0.06187	0.32379
4.50928	0.07526	0.26192
5.21664	0.07108	0.18667
5.92396	0.01076	0.11560
6.63130	0.02409	0.10405
7.33864	0.03191	0.07495
8.04598	0.00953	0.04204
8.75332	0.01071	0.03251
9.46065	0.00616	0.02181
10.16799	0.00543	0.01565
10.87533	0.00478	0.01022
11.58267	0.00222	0.00294
12.29001	0.00112	0.00292
13.00469	0.00014	0.00177
13.82376	0.00005	0.00103
14.64304	0.00041	0.00098
15.46232	0.00043	0.00057
16.28160	0.00004	0.00017
17.10088	0.00003	0.00013
17.92016	0.00006	0.00010
18.73944	0.00001	0.00002
19.55872	0.00000	0.00001
20.37800	0.00000	0.00000
21.19728	0.00000	0.00000
22.01656	0.00000	0.00000

EXACT DISTRIBUTION OF ASU STATISTIC USED.

OBSERVED ASU = 5.21667
 OBSERVED SIGNIFICANCE LEVEL = A1 = 0.18634
 P1 = 2(1-A1) = 0.62732
 P2 = 2(1-P1) = 0.37268
 P3 = 2(1-P2) = 0.62732

Figure VIII.22 EXAX2 output

 HOLCOMBE AND PHIPPS COMPOUND D FRY MORTALITY

GROUP 5		CONTINGENCY TABLE						
TKI	1	ALIVE	DEAD	TOTAL		GRAND TOTAL =	100.000	
	2	2.000	23.000	25.000				
	3	4.000	21.000	25.000				
	4	9.000	16.000	25.000				
TOTAL		ALIVE = 21.000	TOTAL DEAD = 19.000	GRAND TOTAL = 100.000				
EXPECTED FREQUENCIES								
TKI	1	ALIVE	DEAD	TOTAL		GRAND TOTAL =	100.000	
	2	5.250	19.750	25.000				
	3	5.250	19.750	25.000				
	4	5.250	19.750	25.000				
TOTAL		ALIVE = 21.000	TOTAL DEAD = 19.000	GRAND TOTAL = 100.000				

ASYMPTOTIC CHI SQUARED USED.
 OBSERVED CHI SQUARED = 0.44967
 OBSERVED SIGNIFICANCE LEVEL = A1 = 0.09167
 Y1 = -2LN(A1) = 4.17915
 Y2 = E(-2LN(A1)) = 2.00000
 VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.23 EXAX2 output

HOLCOMBE AND PHIPS COMPOUND FRY MORTALITY

GROUP 6

LUNINGENCY TABLE

	1	2	3	4	TOTAL	ALIVE	DEAD	TOTAL	DEAD	TOTAL
IKI						0.0	25.000	25.000	25.000	25.000
IKI						0.0	25.000	25.000	25.000	25.000
IKI						0.0	25.000	25.000	25.000	25.000
IKI						0.0	25.000	25.000	25.000	25.000

TOTAL ALIVE = 0.0 , TOTAL DEAD = 100.000 , GRAND TOTAL = 100.000

TABLE IS DEGENERATE; ROW SUM IS ZERO UK AT LEAST K-1 LUL SUMS ARE ZERO.

OBSERVED SIGNIFICANCE LEVEL = AI = 1.0

TY = -2LN(AI) = 0.0

EY = E(-2LN(AI)) = 0.0

VARY = VARI(-2LN(AI)) = 0.0

6 INDEPENDENT TESTS COMBINED

SY = SUM(TY) = 13.38451

YMU = SUM(EY) = 8.17521

SVART = SUM(VARY) = 17.34129

Z = (SY - SUM(TYMU))/SQRT(SVART/(4*YMU)) = 1.09755

UNDER MU 2 HAS A STANDARD NORMAL DISTRIBUTION

JARVINEN Compound B: BRAYO MORTALITY

GROUP 1

CONTINGENCY TABLE

	1	2	TOTAL
TKI	59.000	12.000	71.000
TKI	49.000	2.000	51.000
TOTAL	ALIVE = 87.000	DEAD = 14.000	GRAND TOTAL = 101.000

EXPECTED FREQUENCIES

CUTOFF = 5.00000

	1	2	TOTAL
TKI	43.931	7.069	51.000
TKI	43.069	6.931	50.000
TOTAL	ALIVE = 87.000	DEAD = 14.000	GRAND TOTAL = 101.000

*** CHI SQUARED HAS BEEN CORRECTED FOR CONTINUITY, SINCE D.F. = 1. ***

ASYMPTOTIC CHI SQUARED USED.
 OBSERVED CHI SQUARED = 0.51208
 OBSERVED SIGNIFICANCE LEVEL = A1 = 0.01071
 Y1 = -2LN(A1) = 7.07234
 EV = E(-2LN(A1)) = 2.00000
 VARY = VAR(-2LN(A1)) = 4.00000

Figure VIII.25 EXAX2 output

 JAVANA COMPANCO A BENSATO MORTALITY

GROUP 2		CONTINGENCY TABLE			
TRT		ALIVE	DEAD	TOTAL	
1		51.000	6.000	57.000	
2		50.000	2.000	52.000	
TOTAL	ALIVE =	101.000	TOTAL DEAD =	8.000	GRAND TOTAL = 109.000
EXPECTED FREQUENCIES					
TRT		ALIVE	DEAD	TOTAL	
1		52.817	4.183	57.000	
2		48.183	3.817	52.000	
TOTAL	ALIVE =	101.000	TOTAL DEAD =	8.000	GRAND TOTAL = 109.000
CUTOFF = 5.00000					

EXACT ASQ DISTRIBUTION

ASQ	PROB	SIG
0.01821	0.258147	1.00000
0.36051	0.24357	0.71851
0.57136	0.20017	0.41444
1.78430	0.12066	0.27477
2.57805	0.08253	0.14812
3.28927	0.03614	0.06429
5.48017	0.02007	0.02640
7.87632	0.00435	0.00633
9.76362	0.00146	0.00198

EXACT DISTRIBUTION OF ASQ STATISTIC USED.
 OBSERVED ASQ = 1.78430
 UNOBSERVED SIGNIFICANCE LEVEL = AI = 0.27477
 VI = -2LN(AI) = 2.58361
 VI = E1-2LN(AI) = 1.522817
 VARI = VARI-2LN(AI) = 3.23301

Figure VIII.26 EXAX2 output

JARVINEN COMPANIES EMBAYED MORTALITY

GROUP 3 CONTINGENCY TABLE

TMT	1	2	ALIVE	DEAD	TOTAL
TMT	1	2	48.000	2.000	50.000
TMT	2	1	43.000	7.000	50.000
TOTAL	ALIVE =	91.000	TOTAL	DEAD =	9.000
					GRAND TOTAL = 100.000

CUTOFF = 5.000000

EXPECTED FREQUENCIES

TMT	1	2	ALIVE	DEAD	TOTAL
TMT	1	2	45.500	4.500	50.000
TMT	2	1	45.500	4.500	50.000
TOTAL	ALIVE =	91.000	TOTAL	DEAD =	9.000
					GRAND TOTAL = 100.000

EXACT KSU DISTRIBUTION

KSU	PROB	SIG
0.12210	0.51303	1.00000
1.09890	0.32747	0.48691
5.05250	0.12865	0.15951
5.98291	0.02822	0.03086
9.89011	0.00263	0.00263

EXACT DISTRIBUTION OF KSU STATISTIC USED.

OBSERVED KSU = 3.05250
OBSERVED SIGNIFICANCE LEVEL = A1 = 0.15951
Y1 = -2.01411
Y2 = 1.17120
VARI = 1.17120
VARI = 2.77600

Figure VIII.27 EXAX2 output

*****X*****

JARVINEJ Compound B EMBRYO MORTALITY

GROUP 4		CONTINGENCY TABLE			
TRT		ALIVE	DEAD	TOTAL	
1		49.000	1.000	50.000	
2		47.000	1.000	48.000	
TOTAL		ALIVE = 96.000	TOTAL DEAD = 2.000	GRAND TOTAL = 98.000	
EXPECTED FREQUENCIES					
TRT		ALIVE	DEAD	TOTAL	
1		48.980	1.020	50.000	
2		47.020	0.980	48.000	
TOTAL		ALIVE = 96.000	TOTAL DEAD = 2.000	GRAND TOTAL = 98.000	
CUTOFF = 5.000000					

EXACT ASU DISTRIBUTION

ASU	PROB	SIG
0.00085	0.50494	1.00000
1.96660	0.25773	0.45506
2.14674	0.23732	0.23732

EXACT DISTRIBUTION OF ASU STATISTIC USED.

OBSERVED ASU = 0.00085

OBSERVED SIGNIFICANCE LEVEL = AI = 1.00000

YY = -2LN(AI) = 0.00000

EY = E(-2LN(AI)) = 1.64511

VARY = VAR(-2LN(AI)) = 1.38125

Figure VIII.28 EXAX2 output

CONTINGENCY TABLE					
	ALIVE	DEAD	TOTAL		
TMT 1	78.000	2.000	80.000		
TMT 2	51.000	2.000	53.000		
TOTAL	129.000	4.000	133.000	GRAND TOTAL =	133.000
				CUTOFF =	5.000000

EXPECTED FREQUENCIES					
	ALIVE	DEAD	TOTAL		
TMT 1	79.071	2.929	82.000		
TMT 2	49.929	1.071	51.000		
TOTAL	129.000	4.000	133.000	GRAND TOTAL =	133.000

EXACT AND DISTRIBUTION

XSD	PKGB	SIG
0.00000	0.32109	1.000000
0.00730	0.24127	0.678850
0.07812	0.23181	0.437804
0.17830	0.09272	0.205233
0.30093	0.08559	0.112511
0.44351	0.01421	0.045924
0.48001	0.01265	0.041265

EXACT DISTRIBUTION OF ASD STATISTIC USED.

OBSERVED ASD	0.48001
OBSERVED SIGNIFICANCE LEVEL	0.437804
TT = -ZLN(A1)	1.05548
ET = E(-ZLN(A1))	1.45272
VARY = VAK(-ZLN(A1))	2.59750

 TAGVAREO COMPANY A BINARY MORTALITY

GROUP 6 CONTINGENCY TABLE

TKT	1	2	ALIVE	DEAD	TOTAL
TKT	1	2	44.000	7.000	51.000
TOTAL	ALIVE =	93.000	TOTAL	DEAD =	8.000
			GRAND TOTAL =		101.000

CUTOFF = 5.000000

EXPECTED FREQUENCIES

TKT	1	2	ALIVE	DEAD	TOTAL
TKT	1	2	40.040	5.960	50.000
TOTAL	ALIVE =	93.000	TOTAL	DEAD =	8.000
			GRAND TOTAL =		101.000

EXACT KSQ DISTRIBUTION

KSQ	PROB	SIG
0.00000	0.00000	1.00000
0.00000	0.00000	0.71522
0.00000	0.00000	0.48478
0.00000	0.00000	0.28471
0.00000	0.00000	0.18910
0.00000	0.00000	0.10025
0.00000	0.00000	0.05466
0.00000	0.00000	0.03101
0.00000	0.00000	0.00501
0.00000	0.00000	0.00200

EXACT DISTRIBUTION OF KSQ STATISTIC USED

OBSERVED KSQ = 4.72938
 OBSERVED SIGNIF. LEVEL = .00000
 #1 = 2.00000
 #2 = 2.00000
 #3 = 2.00000
 #4 = 2.00000
 #5 = 2.00000
 #6 = 2.00000
 #7 = 2.00000
 #8 = 2.00000
 #9 = 2.00000
 #10 = 2.00000
 #11 = 2.00000
 #12 = 2.00000
 #13 = 2.00000
 #14 = 2.00000
 #15 = 2.00000
 #16 = 2.00000
 #17 = 2.00000
 #18 = 2.00000
 #19 = 2.00000
 #20 = 2.00000
 #21 = 2.00000
 #22 = 2.00000
 #23 = 2.00000
 #24 = 2.00000
 #25 = 2.00000
 #26 = 2.00000
 #27 = 2.00000
 #28 = 2.00000
 #29 = 2.00000
 #30 = 2.00000
 #31 = 2.00000
 #32 = 2.00000
 #33 = 2.00000
 #34 = 2.00000
 #35 = 2.00000
 #36 = 2.00000
 #37 = 2.00000
 #38 = 2.00000
 #39 = 2.00000
 #40 = 2.00000
 #41 = 2.00000
 #42 = 2.00000
 #43 = 2.00000
 #44 = 2.00000
 #45 = 2.00000
 #46 = 2.00000
 #47 = 2.00000
 #48 = 2.00000
 #49 = 2.00000
 #50 = 2.00000
 #51 = 2.00000
 #52 = 2.00000
 #53 = 2.00000
 #54 = 2.00000
 #55 = 2.00000
 #56 = 2.00000
 #57 = 2.00000
 #58 = 2.00000
 #59 = 2.00000
 #60 = 2.00000
 #61 = 2.00000
 #62 = 2.00000
 #63 = 2.00000
 #64 = 2.00000
 #65 = 2.00000
 #66 = 2.00000
 #67 = 2.00000
 #68 = 2.00000
 #69 = 2.00000
 #70 = 2.00000
 #71 = 2.00000
 #72 = 2.00000
 #73 = 2.00000
 #74 = 2.00000
 #75 = 2.00000
 #76 = 2.00000
 #77 = 2.00000
 #78 = 2.00000
 #79 = 2.00000
 #80 = 2.00000
 #81 = 2.00000
 #82 = 2.00000
 #83 = 2.00000
 #84 = 2.00000
 #85 = 2.00000
 #86 = 2.00000
 #87 = 2.00000
 #88 = 2.00000
 #89 = 2.00000
 #90 = 2.00000
 #91 = 2.00000
 #92 = 2.00000
 #93 = 2.00000
 #94 = 2.00000
 #95 = 2.00000
 #96 = 2.00000
 #97 = 2.00000
 #98 = 2.00000
 #99 = 2.00000
 #100 = 2.00000

INDEPENDENT TESTS COMBINED

ST = SUM(XI) = 2.00000
 TAU = SUM(XI) = 8.71810
 VARI = SUM(VARI) = 17.51398
 Z = (TAU - ST) / (SUM(VARI) / SUM(XI)) = 2.54488
 UNDER NO L HAD A STANDARD NORMAL DISTRIBUTION

Figure VIII.30 EXAX2 output

ROUND 6 | FRY MORTALITY

[illegible]

Figure VIII.31 EXAX2 output

JARVINEN COMPOUND 8 FRY MORTALITY

GROUP 2

CONTINGENCY TABLE

		ALIVE	DEAD	TOTAL
IR1	1	15,000	0.0	15,000
IR1	2	14,000	0.0	14,000
TOTAL	ALIVE	29,000	TOTAL DEAD	0.0
				GRAND TOTAL = 29,000

TABLE IS ORDERED BY ROW SUM, IS ZERO OR AT LEAST K=1 COL SUMS ARE ZERO.

CHS (AVD STONE) CANCE LEVEL = A1 = 1.0

YY = 2LN(A1) = 0.0

XY = 1(-2LN(A1)) = 0.0

VARY = VAR(-2LN(A1)) = 0.0

Figure VIII.32 EXAX2 output

 JARVINEN COMPANY B FRY MORTALITY

GROUP 3

CONTINGENCY TABLE

IR1	1	2	ALIVE	DEAD	TOTAL
IR1	1	2	15.000	0.0	15.000
IR1	2	1	15.000	0.0	15.000
TOTAL	ALIVE =	30.000	TOTAL	DEAD =	0.0
					GRAND TOTAL = 30.000

TABLE IS UNCONDITIONED. ROW SUM IS ZERO OR AT LEAST K-1 COL SUMS ARE ZERO.
 OR: $VAR = 0$ (SIGNIFICANCE LEVEL = $\alpha = 0.0$)
 $YY = -2LN(A1) = 0.0$
 $XY = E(-2LN(A1)) = 0.0$
 $VARY = VAR(-2LN(A1)) = 0.0$

Figure VIII.33 EXAX2 output

VARYING COMPOUND B FAY MORTALITY

CONTINGENCY TABLE

GROUP 4

	1	2	ALIVE	DEAD	TOTAL
1	15.000	0.0	15.000	0.0	15.000
2	15.000	0.0	15.000	0.0	15.000
TOTAL	30.000	0.0	30.000	0.0	30.000

TABLE IS DEAD NEARLY ZERO FROM LEAST K-1 COL SUMS ARE ZERO.

TESTED SIGNIFICANCE LEVEL = $\alpha = 0.0$

$Y = -2LN(A1) = 0.0$

$Y = -2LN(A1) = 0.0$

$VAR = VAR(-2LN(A1)) = 0.0$

Figure VIII.34 EXAX2 output

JARVINEN Compound B FAY MORTALITY

GROUP 5

CONTINGENCY TABLE

	1	2	TOTAL
TRF	1	7.000	7.000
TRF	2	9.000	9.000
TOTAL	16.000	16.000	32.000

TOTAL ALIVE = 16.000 , TOTAL DEAD = 16.000 , GRAND TOTAL = 32.000

EXPLODED FREQUENCY

CUTOFF = 5.00000

	1	2	ALIVE	DEAD	TOTAL
TRF	1	7.000	7.000	0.000	7.000
TRF	2	9.000	0.000	9.000	9.000
TOTAL	16.000	16.000	7.000	9.000	16.000

TOTAL ALIVE = 16.000 , TOTAL DEAD = 16.000 , GRAND TOTAL = 32.000

*** CHI SQUARED HAS BEEN ADJUSTED FOR CONTINUITY, SINCE D.F. = 1. ***

ASYMPTOTIC CHI SQUARED USED.

OBSERVED CHI SQUARED = 0.12353
 OBSERVED SIGNIFICANT LEVEL = 0.71435
 P-VALUE = 0.71435
 VARIANCE = 0.00000
 VARIANCE = 0.00000

Figure VIII.35 EXAX2 output

COMPOUND 8 FAY MORTALITY

CONFIDENTIAL

	ALIVE	DEAD	TOTAL
1	9.0	15.000	15.000
2	0.0	15.000	15.000
TOTAL	9.0	30.000	30.000

[illegible]
$$YY = -2L^2(x) = -2L^2(0) = -2$$
$$Y = (Y_1, \dots, Y_n)' = (Y_1, \dots, Y_{n-1}, Y_n)'$$
$$\text{VAR} = \text{VAR}(-\text{LN}(AT)) = 1.0$$

6 INDEPENDENT TESTS COMBINED

$$Y = T(Y)$$
$$Y_{MU} = \sum Y_{ij} = 2,000,000$$
$$\text{CVAFY} = \text{SUM(VAFY)} = 4.00000$$
$$Z = (\sqrt{RT}(\beta Y) - \sqrt{RT}(\gamma M^2)) / \sqrt{RT}(\beta Y + \gamma M^2) = -0.84013$$

UNDER THE STANDARD NORMAL DISTRIBUTION

APPENDIX AVIII.2^{*} DESCRIPTION AND INSTRUCTIONS FOR USE OF EXAX2
COMPUTER PROGRAM

^{*}Appendix AVIII.2 is an appendix for Section .III.

EXAX2--A COMPUTER PROGRAM TO
COMPARE BINOMIAL PROPORTIONS*

Program Description and Card Input Information

Paul I. Feder and Susan A. Willavize
Department of Statistics
The Ohio State University

September 12, 1980

* This work was performed with the support of the U.S. Army Medical Bioengineering Research and Development Laboratory, Frederick, Maryland under Contract DAMD17-79-C-9150 at The Ohio State University.

EXAX2--A Computer Program to Compare Binomial Proportions

I. Introduction

Dichotomous data arise in many fields of application. In fish toxicology experiments it is of interest whether an embryo hatches into a live fry, whether the fry has survived 27 days post hatching, whether the surviving fry are normal or abnormal. In toxicology experiments on higher animals, such as mammals, the presence or absence of specific types of tumors or deformities is of importance. Such responses naturally give rise to 0-1 or success-failure type data. Such 0-1 data are encountered also in many other fields. For example in industrial applications it is noted whether or not a unit meets design specifications or whether or not a unit lasts beyond the warranty period. In sociological applications it is of interest to note whether an individual exhibits certain behavior patterns, has specific opinions, etc. A myriad of additional applications could be cited.

Part of the analysis of success-failure data involves estimating the probabilities of "success" within various groups and comparing these probabilities across groups. EXAX2 is a computational tool to assist in carrying out such comparisons. Since the need for the program was motivated by problems arising in aquatic toxicology, the remainder of the section centers around such applications.

In toxicity tests on fish or daphnids, a number of test concentrations are run along with one or more control groups. Within each concentration group (treatment or control) several tanks or beakers are run, each chamber containing a number of the organisms under study. Within each chamber the numbers of embryos, numbers of fry hatched live or normal, numbers of fry surviving or normal at the conclusion of the experiment are recorded. It is desired to compare the proportions of live or normal embryos or fry in the various treatment groups with corresponding proportions in the control groups.

A preliminary inference of importance is to test the response proportions among the tanks or beakers within each group for homogeneity. If there is no evidence of tank to tank heterogeneity within concentration groups, the data can be pooled across tanks within groups and further analyses carried out based on binomial theory. However if evidence exists of tank to tank heterogeneity within groups, subsequent analyses must be adjusted to reflect this, either by adjusting the model or the data or by carrying out analyses on a per tank basis.

EXAX2 carries out tests of homogeneity of tanks within groups based on the chi square statistic. If the expected response frequencies are "large enough" the distribution of the test statistic is approximated by large sample chi square theory. If the expected response frequencies are not large enough for asymptotic theory to be applicable, the test statistic is evaluated based on its exact small sample distribution, derived from the exact small sample distribution of the contingency table, conditional on the margins (March, 1972). Individual tests of homogeneity within groups are combined by means of Fisher's method (Littell and Folks, 1971, 1973) to obtain an overall test of homogeneity.

EXAX2 also has the capability to test for heterogeneity of response rates across treatment groups based on responses pooled within groups. Either the exact small sample or approximate large sample distribution of the chi square statistic is utilized. Heterogeneity among tanks within groups can be accounted for either by modifying the chi square statistic by a heterogeneity factor (Finney, 1971) or by modifying the data to "effective frequencies" to reflect within tank correlation.

EXAX2 can calculate exact confidence intervals on the odds ratio of a treatment group response rate as compared with a control group response rate, based on the small distribution of Fisher's exact test for 2x2 tables (Thomas, 1971). The confidence interval calculations are based on responses pooled across tanks within

The response frequencies would need to be modified to "effective frequencies"

to reflect within tank correlation.

Section II discusses program organization and capabilities and provides a more detailed description of the program's procedures. Section III contains detailed instructions for card input.

II Program Organization and Capabilities

Suppose that the aquatic toxicity test consists of N treatment groups (including both test and control groups) and K tanks or beakers per group. (EXAX2 can handle different values of K for each group, however we assume a single K value for notational convenience.) Thus the responses within each treatment group can be summarized as a $2 \times K$ contingency table. The rows represent the response category (e.g. dead, live or abnormal, normal, etc.) Each column represents the responses from an individual tank. We wish to compare response probabilities across columns.

It is first necessary to specify the entries in the tables. A $2 \times K$ contingency table can be specified as a $3 \times (K+1)$ matrix partitioned as follows:

	Col 1	Col 2	...	Col K	Row Totals
Row 1	$X(1, 1)$	$X(1, 2)$...	$X(1, K)$	$R(1)$
Row 2	$X(2, 1)$	$X(2, 2)$...	$X(2, K)$	$R(2)$
Col Total	$C(1)$	$C(2)$...	$C(K)$	XN

where the $X(I, J)$ comprise the "body" of the table; $R(1)$ and $R(2)$ are row totals; $C(1)$, $C(2)$, ..., $C(K)$ are the column totals; and XN is the grand total.

The information in the first K columns of each matrix is inputted one column at a time either by specifying the body of the table ($X(1, 1)$, $X(2, 1)$, ..., $X(1, K)$, $X(2, K)$) or the column totals and the first row ($C(1)$, $X(1, 1)$, ..., $C(K)$, $X(1, K)$), or the column totals and the second row ($C(1)$, $X(2, 1)$, ..., $C(K)$, $X(2, K)$) (See section III, instructions for card input). The remaining elements are computed and the complete matrix is printed. A single matrix or several matrices must be inputted, depending on the purpose of the analysis.

Each inputted matrix is first examined to detect and adjust for the following conditions -- if they exist:

1.) If any column totals are zero those columns are deleted, K is reduced accordingly, and the new matrix is printed.

2.) If any row total is zero or if only one column total is nonzero, then that table is the only possible table with the given row and column totals. As such it is defined to be degenerate and the observed level of significance (for future heterogeneity tests) is set to 1. A message to this effect is printed.

3.) Steps 1 and 2 above are repeated for each succeeding inputted contingency table.

We first consider tests of homogeneity within treatment groups and later we will discuss tests across treatment groups.

If a table is not degenerate, the expected cell frequencies ($EX(I, J)$) are calculated as $EX(I, J) = R(I) C(J) / XN$ ($I = 1, 2; J = 1, \dots, K$) and XSQ is calculated as $XSQ = \sum_{J=1}^K \sum_{I=1}^2 (X(I, J) - EX(I, J))^2 / EX(I, J)$. If $K = 2$ a correction for continuity is applied to improve the convergence of the distribution of XSQ to its asymptotic chi square form. Namely if $K = 2$

$$XSQ = \sum_{J=1}^2 \sum_{I=1}^2 (|X(I, J) - EX(I, J)| - 1/2)^2 / EX(I, J).$$

The table of expected frequencies and the user specified cutoff value, CUTOFF (for what constitutes a "large" expected frequency within each cell), are printed.

The table of expected frequencies is then examined to see if any of the expected frequencies are less than CUTOFF, e.g. 5. If not XSQ is considered to be asymptotically distributed as χ^2 with $K-1$ degrees of freedom, and its significance level (A_i) is calculated based on chi square theory. The observed chi square value and its significance level (A_i) are printed. If one or more of the expected frequencies is less than CUTOFF, the exact distribution of the XSQ statistic is calculated. This is done by enumerating all possible tables with the given row

and column totals (algorithm due to Boulton and Wallace, 1973) and their associated chi square values. Under the assumption of homogeneity of response probabilities across columns the probability of each possible table, conditional on the row and column margins fixed, is (March, 1972)

$$\frac{R(1)! R(2)! \dots R(J)!}{XN! \prod_{I,J} X(I, J)!} \prod_{I,J} K$$

From this exact distribution over possible tables the exact distribution of XSQ is derived. Based on this derived distribution, the significance level (A_i) of the XSQ value associated with the observed table (XSQOBS) is calculated as the probability of a XSQ value greater than or equal to XSQOBS. The observed XSQ value and significance level (A_i) are printed and optionally (see instructions for card input) the exact XSQ distribution is printed. The entire process is then repeated on the next contingency table.

After tests of homogeneity (asymptotic or exact) have been carried out on each treatment group, the significance levels A_1, A_2, \dots, A_n summarize the results of the independent tests of the homogeneity on each table. To obtain an overall significance level these independent A_i 's are combined as follows:

For groups where the distribution of XSQ has been approximated by its asymptotic chi square form, the null distribution of A_i is approximately uniform (0, 1). Thus $Y_i \equiv -2 \ln (A_i)$ has an approximate chi square distribution with 2 degrees of freedom, mean $E(Y_i) \equiv EY_i \equiv E(-2 \ln (A_i)) = 2$, and variance $\text{Var} (Y_i) \equiv \text{VARY}_i \equiv \text{Var} (-2 \ln (A_i)) = 4$.

For groups where the exact small sample distribution of XSQ has been used, A_i and $Y_i \equiv -2 \ln (A_i)$ have discrete null distributions derived from the null distribution of the contingency table. The mean $E(Y_i) \equiv EY_i$ and variance $\text{Var} (Y_i) \equiv \text{VARY}_i$ are calculated from the exact distribution of $Y_i = -2 \ln (A_i)$.

In either case (i.e. exact or asymptotic) the values of Y_i , EY_i , and $VARY_i$ are calculated and printed along with the other results for each table. The test statistic for the overall test of tank to tank homogeneity within groups is

$$Z = \frac{\sqrt{EY_i} - \sqrt{\sum EY_i}}{\sqrt{\sum VARY_i / (4 \sum EY_i)}}$$

It is calculated and printed at the end of the output. Under the null hypothesis that the tables are all homogeneous, Z has an approximate standard normal distribution. The null hypothesis is rejected for large values of Z.

We now consider additional applications of EXAX2. The program has the capability to carry out chi square tests of homogeneity across treatment groups and to construct exact confidence intervals on the odds ratios of treatment groups compared with the control. These applications are discussed in turn, beginning with tests of homogeneity across groups.

If preliminary tests do not reveal heterogeneity among tanks within treatment groups, then it is appropriate to sum the observed frequencies in the individual tanks within each treatment group. This results in a new $2 \times N$ contingency table which can be tested for homogeneity across treatment groups. EXAX2 can perform the appropriate summing within groups and then proceed with the chi square test across groups, based either on exact or asymptotic theory, as discussed above. Since just one contingency table is involved in this application, the Z statistic is not computed.

EXAX2 has the capability to compare, on a pairwise basis, the odds ($p/(1-p)$) within each treatment group to the odds in a user specified group; e.g. the control group. Using an algorithm given by Thomas (1971) an exact confidence interval is computed for each odds ratio (one per treatment group). The user specifies both the upper and lower alpha levels, thus permitting either one sided or two sided confidence intervals.

To illustrate how this works consider the frequencies for treatment 1 (the control group) and treatment T as forming a 2x2 contingency table:

	Col 1	Col T	Row Totals
Row 1	X(1, 1)	X(1, T)	Y(1)
Row 2	X(2, 1)	X(2, T)	Y(2)
Col Totals	C(1)	C(T)	YN

Y(1), Y(2), YN designate respectively the two row totals and the grand total of this new table. The odds ratio PSI is defined as $PSI = (P_1 Q_T) / (P_T Q_1)$ where P_1, P_T are the category 2 probabilities (i.e. "success") within treatment groups 1, T respectively and $Q_1 \equiv 1 - P_1, Q_T \equiv 1 - P_T$ are the category 1 probabilities within treatment groups 1, T respectively.

We estimate these quantities by

$$\hat{P}_1 = X(2, 1) / C(1)$$

$$\hat{P}_T = X(2, T) / C(T)$$

$$\hat{Q}_1 = X(1, 1) / C(1) \equiv 1 - \hat{P}_1$$

$$\hat{Q}_T = X(1, T) / C(T) \equiv 1 - \hat{P}_T$$

$$\hat{PSI} = (X(2, 1) X(1, T)) / (X(1, 1) X(2, T)).$$

Thomas' algorithm is an iterative technique for finding upper and lower confidence bounds on PSI. It is based on the noncentral distribution of Fisher's exact test statistic.

III Instructions for Card Input

In this section we present detailed instructions for card input to EXAX2.

The card input consists of 5-12 Program Information Cards followed by the Data Cards and the Alpha Card. The Program Information Cards are, in order: one Input Option Card, one Parameter Card, one Format Card, one Header Card, one Labels Card, and from one to six Title Cards. These cards must be punched as described below:

1. INPUT OPTION CARD This card should have a 1 in card column 1 if the

subsequent data cards represent K tanks per treatment group and the tank data is to be pooled (summed) within each treatment group to produce one 2XN contingency table which will then be analyzed. There should be a 2 in card column 1 if no pooling is to occur. In this case, the data will be tested for homogeneity among tanks within treatment groups.

2. PARAMETER CARD

<u>Card Cols.</u>	<u>Description</u>
1-5	K = the number of columns in each contingency table (a right justified integer) (K is the number of tanks per treatment group.) K should be less than or equal to 12 if the number on the Input Option Card is 2.
6-15	CUTOFF = the smallest expected cell frequency with which the use of the asymptotic chi square approximation is permitted. If one or more expected cell frequencies are smaller than CUTOFF then the exact small sample distribution of XSQ is used. (CUTOFF is a real number with decimal point.)
16-19	blanks
20	NTITLE = the number of Title Cards used (an integer from 1 to 6)
21-25	N = the number of 2XK contingency tables, i.e. the number of groups (both treatment and control.) If the number on the Input Option Card is 1, then N should be less than or equal to 12.
26-29	blanks
30	IOPT = 1, if the exact distribution of XSQ is to be printed = 0, otherwise.
31-34	blanks
35	IDATA = 1, if data card input is of the form C(I), X(1, I) (see the description of the Data Cards given below for an explanation of this notation) = 2 if data card input is C(I), X(2, I) = 3 if data card input is X(1, I), X(2, I)

<u>Card Cols.</u>	<u>Description</u>
-------------------	--------------------

40

IC = group number corresponding to group (control or treatment) to which each other treatment group is to be compared when calculating confidence intervals on the odds ratios (e.g. the number of the control group).

It should be noted with respect to the specification of K in columns 1-5 of the parameter card that some groups may have L<K tanks. The associated contingency tables for these groups thus have L<K columns. These tables must be augmented with K-L columns of zeros by inputting K-L Data Cards containing zeros. The program will later delete these dummy columns and perform its computations based only on the original L columns.

3. FORMAT CARD

This card contains the format in which the subsequent data cards have been punched, e.g. (F5.0, F5.0), (see the description of the data cards given below).

Note: No I or A formats may be used! All 80 card columns may be used. The format statement may be positioned anywhere on the card and must include the parentheses (see any FORTRAN text for an explanation of **formats**).

4. HEADER CARD

This contains a special heading (it may be blanks) which will be printed above the output for each contingency table. Any or all of the 80 columns may be used. (e.g. Embryo Mortality)

5. LABELS CARD

<u>Card Cols.</u>	<u>Description</u>
-------------------	--------------------

1-2

blanks

3-10

A label for the first row of the contingency table

11-12

blanks

13-20

A label for the second row of the contingency table (e.g. "live", "dead")

6. TITLE CARDS

These cards contain titles which will be printed above the output for each contingency table. Any or all of the 80 columns on each card may be used. The number of Title Cards must agree with the entry NTITLE on the Parameter Card. If no titles are desired there must be a blank card here and NTITLE must equal 1.

7. DATA CARDS

An observed $2 \times K$ contingency table can be specified as a $3 \times (K+1)$ matrix partitioned as follows:

	Col 1	Col 2	...	Col K	Row Totals
Row 1	X(1, 1)	X(1, 2)	...	X(1, K)	R(1)
Row 2	X(2, 1)	X(2, 2)	...	X(2, K)	R(2)
Col. Totals	C(1)	C(2)	...	C(K)	XN

The $X(I, J)$ comprise the "body" of the table; $R(1)$ and $R(2)$ are row totals; $C(1)$, $C(2)$, ..., $C(K)$ are the column totals; and XN is the grand total. Each Data Card contains the content of one of the first K columns of this matrix. This information can be specified in one of three ways. If IDATA = 1 (in card column 35 of the Parameter Card), the I -th Data Card must contain $C(I)$, $X(1, I)$. If IDATA = 2, the I -th Data Card must contain $C(I)$, $X(2, I)$. If IDATA = 3, the I -th Data Card must contain $X(1, I)$, $X(2, I)$. EXAX2 will compute the other matrix elements. The format given on the second Program Information Card specifies the format under which this information will be read.

The K Data Cards for the first contingency table are followed by the K Data Cards for the second contingency table and so on until all N contingency tables have been given. There will thus be $N \times K$ Data Cards in all. The form of the data for each contingency table must be consistent with the entries on the Program Information Cards.

8. ALPHA CARD

<u>Card Cols</u>	<u>Description</u>
1-5	ALPHAL = the lower alpha level for the odds ratio confidence intervals
6-10	ALPHAU = the upper alpha level for the odds ratio confidence intervals.

These must be decimal numbers less than or equal to one and their sum must be less than or equal to one.

If the data consist of just one tank per group or if EXAX2 is instructed to pool data across tanks within groups, then confidence intervals on the odds ratios are automatically computed. If the data consist of more than one tank per group (i.e. $K > 1$) and EXAX2 is instructed not to pool across tanks within groups then confidence intervals are not computed and this card may be omitted.

The input cards are placed at the end of the program deck between the //GØ.SYSIN DD * card and the /* card. The example below illustrates the input stream.

1

4 5.0 5 5 1 1 1
 (15X,F5.0,F5.0)

EMBRYO MORTALITY

ALIVE DEAD

INVESTIGATOR: HULCOMBE & PHIPPS (06), TEST: D (03)
 DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
 SIX LEVELS: 1(CONTROL), 2(LOWEST), ..., 6(HIGHEST); 4 REPS EA (A,B,C,D)
 # EMBRYOS TESTED, # LIVE AFTER HATCH, # NORMAL FRY AFTER HATCH
 # FRY TESTED, # ALIVE AT END, # NORMAL AT END

1 A 1 10603	50	33	33	25	23	23
1 B 1 10603	50	29	28	25	23	23
1 C 1 10603	50	38	36	25	24	24
1 D 1 10603	50	30	30	25	24	24
2 A 1 10603	50	31	31	25	22	22
2 B 1 10603	50	36	35	25	24	24
2 C 1 10603	50	34	33	25	23	23
2 D 1 10603	50	29	28	25	23	23
3 A 1 10603	50	35	35	25	23	23
3 B 1 10603	50	39	39	25	25	25
3 C 1 10603	50	26	26	25	20	20
3 D 1 10603	50	38	36	25	24	24
4 A 1 10603	50	30	30	25	21	21
4 B 1 10603	50	29	29	25	20	20
4 C 1 10603	50	27	25	25	21	21
4 D 1 10603	50	36	35	25	25	25
5 A 1 10603	50	32	31	25	02	02
5 B 1 10603	50	31	27	25	04	04
5 C 1 10603	50	36	35	25	09	09
5 D 1 10603	50	34	33	25	06	06
6 A 1 10603	50	36	00	25	00	00
6 B 1 10603	50	34	00	25	00	00
6 C 1 10603	50	25	00	25	00	00
6 D 1 10603	50	29	00	25	00	00

0.0250.025

In this example the tank data will be pooled within treatments and 95% confidence intervals (ALPHA = 0.025, ALPHAU = 0.025) will be computed on the odds ratios between treatment one and every other treatment. There are six treatments with 4 tanks each. The exact distribution of XSQ will be printed if it is used in the homogeneity test; i.e., if any of the expected frequencies is less than 5.0 (CUTOFF). There are five title cards preceeding the data cards from which the total number of embryos and the number alive will be read.

IV Program Limitations

EXAX2 has limitations of time and space. The space limitations are due to the dimensioned size of various arrays. When pooling tanks within treatments, N must be less than or equal to 12 and when not pooling tanks, K must be less than or equal to 12. This limit can be raised by increasing the sizes of the arrays in the following table:

<u>Program Location</u>	<u>Arrays</u>
MAIN program	X, C, EX, ICHECK
SUBROUTINE TABLE	X, C
SUBROUTINE INPUT	X, C
SUBROUTINE INPUT1	X, C
SUBROUTINE INPUT2	X, C
SUBROUTINE EXAX	Y, RX, XSQT, D, V, U, DLLM

Depending upon the number of tables (with the given margins) that are enumerated in generating the exact distribution of XSQ, it may also be necessary to increase the sizes of the arrays SIG, XSQ, PROB and IPOINT in SUBROUTINE EXAX. Also in FUNCTION FUN (a part of Thomas' confidence interval algorithm) there is a machine-dependent constant (DPC) which should be set in the DATA statement to the largest real number the machine can hold.

In some cases the CPU time required by the program may be quite large. No systematic study has been performed to determine when this is so, but the following two examples may be of interest:

1.) For a 2x10 table with column margins all equal to 50 and row margins of 46 and 454, the smallest expected cell frequency is 4.6. If CUTOFF = 5.0, the exact distribution of XSQ will be generated requiring the enumeration of over 435,000 tables and a CPU time of over 10 minutes on the AMDAHL 470 V6 computer.

2.) For a 2x6 table with column margins all equal to 60 and row margins of 333 and 27, the smallest expected frequency is 4.5. If CUTOFF = 5.0, the exact distribution of XSQ will be generated requiring a CPU time of over 50 seconds.

It should also be noted that, as Agresti and Wackerly (1977) point out, generally for a fixed grand total the number of tables enumerated (and hence the CPU time) in generating the exact distribution is much higher when the column margins are equal than when they are unequal.

References

1. Agresti, A. and Wackerly, D. (1977). Some Exact Conditional Tests of Independence for $R \times C$ Classification Tables. Psychometrika, 42, pp. 111-125.
2. Boulton, D. M. and Wallace, C. S. (1973). Occupancy of a Rectangular Array. Computer Journal, 16, pp. 57-63.
3. Finney, D. J. (1971). Probit Analysis, 3rd ed., Cambridge University Press.
4. Littell, R. C. and Folks, J. L. (1971, 1973) Asymptotic Optimality of Fisher's Method of Combining Independent Tests, I and II, Journal of the Amer. Statist Assn, 66, pp. 802-806 and 68, pp. 193-194.
5. March, D. L. (1972). Algorithm 434: Exact probabilities for $R \times C$ contingency tables. Comm. of ACM, 15, pp. 991-992.
6. Thomas, D. G. (1971). Exact Confidence Limits for the Odds Ratio in a 2×2 Table. Applied Statistics, 20, pp. 105-110.

AD-A134 389

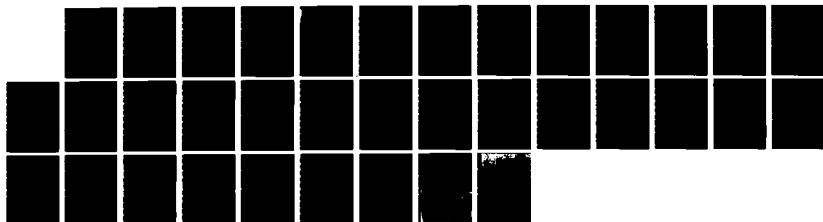
DESIGN AND ANALYSIS OF CHRONIC AQUATIC TESTS OF
TOXICITY(U) OHIO STATE UNIV RESEARCH FOUNDATION
COLUMBUS P I FEDER ET AL. MAY 81 DAMD17-79-C-9150

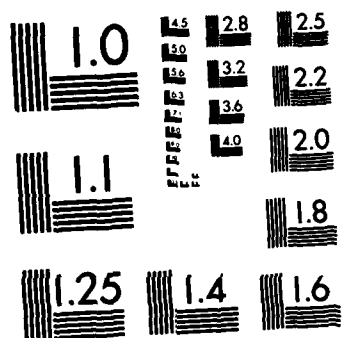
5/5

UNCLASSIFIED

F/G 6/20

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

APPENDIX AX*. THEORETICAL BASES OF SUGGESTED OUTLIER DETECTION TRANSFORMATIONS

In this appendix we discuss the motivations and theoretical bases underlying the outlier detection procedures that are illustrated in the body of the section.

Suppose that the data originate from I treatment groups, J tanks per group. Consider the i-th group. Let X_{ij} , N_{ij} denote the number of responses and the total number of fish in the j-th tank, $j = 1, \dots, J$, and let \hat{p}_i denote the pooled estimated response rate ($\hat{p}_i = \sum_j X_{ij} / \sum_j N_{ij}$), $\hat{q}_i = 1 - \hat{p}_i$.

In subsequent discussion we omit the subscript i for notational convenience and so these quantities are denoted as \hat{p} , \hat{q} , X_j , N_j , respectively. Let $N \equiv \sum_j N_j$.

Consider the chi square statistic for testing for tank to tank heterogeneity within groups

$$\chi^2 \equiv \sum_{j=1}^J \frac{(X_j - N_j \hat{p})^2}{N_j \hat{p} \hat{q}}$$

This statistic is distributed as chi square with J-1 d.f. under the null hypothesis of no tank to tank heterogeneity. For the purpose of detecting outlying responses we consider three cases:

- Case 1: All expected frequencies within the group are greater than a specific cutoff, e.g. 5.
- Case 2: $\hat{p} < 0.1$ or $\hat{p} > 0.9$.
- Case 3: $0.1 \leq \hat{p} \leq 0.9$ and one or more expected frequencies of responses is less than the cutoff, e.g. 5.

We suggest somewhat different transformation in each case.

Case 1. All the expected frequencies within the group exceed the cutoff

Consider the individual terms in the chi square test statistic, $(X_j - N_j \hat{p}) / [N_j \hat{p} \hat{q}]^{1/2}$. Assume that the weights in the denominator are "correct" and "fixed". The quantities X_j , \hat{p} in the numerator are correlated since \hat{p} includes X_j . It can be shown that the variance of

* Appendix AX is an appendix for Section X.

$(X_j - N_j \hat{p}) / [N_j \hat{p} \hat{q}]^{1/2}$ is $1 - (N_j/N)$. If all the N_j 's are equal, this variance is $1 - 1/J$. If all the expected frequencies in the table exceed 5, as is the case with the Defoe compound C embryo mortality data, then the quantities

$$(1 - N_j/N)^{-1/2} \left(\frac{X_j - N_j \hat{p}}{[N_j \hat{p} \hat{q}]^{1/2}} \right) \quad j = 1, \dots, J$$

can be treated as having an approximate standard normal distribution. Graphical and numerical outlier detection procedures are based on these standardized ratios. We pool them across groups and plot the resulting IJ values on normal probability paper. However, for the purpose of formal inference we account approximately for the correlation among terms within groups (approximately $-1/(J - 1)$) by treating the J values within each group as if they were $J - 1$ independent values. This adjustment of course has the most impact when $J = 2$. The normality assumption might be enhanced by first carrying out an arc sine variance stabilizing transformation.

Case 2. Expected response probability within the group is small
e.g. $\hat{p} < 0.1$

Case 2 is also applicable to the situation when $\hat{p} > .9$, by considering the complementary response.

The distribution of X_j can be approximated by a Poisson distribution with mean $\lambda_j \equiv N_j \hat{p}$. The variance stabilizing transformation in the Poisson case is well known to be the square root transformation. In particular, $2(X_j^{1/2} - \lambda_j^{1/2})$ has an approximate standard normal distribution. We estimate λ_j by $N_j \hat{p} \equiv \hat{\lambda}_j$. Now $X_j, \hat{\lambda}_j$ are positively correlated since $\hat{\lambda}_j$ includes X_j . It can be shown by a Taylor expansion argument that the variance of $2(X_j^{1/2} - \hat{\lambda}_j^{1/2})$ is approximately $(1 - N_j/N)$. If all the N_j 's are equal, this variance is $1 - 1/J$.

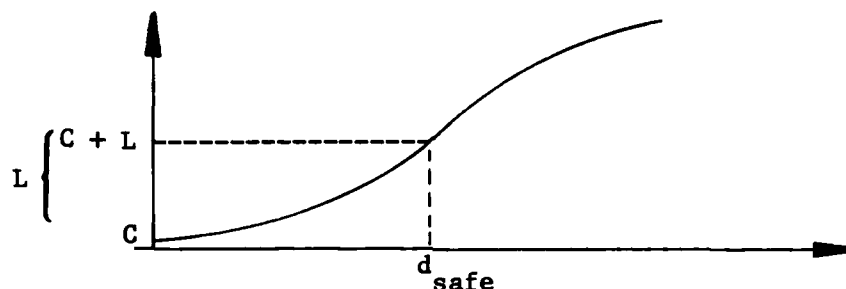
Thus if $\hat{p} < 0.1$ (or if $\hat{q} < 0.1$), the quantities $(1 - N_j/N)^{1/2} \times 2(X_j^{1/2} - (N_j \hat{p})^{1/2})$, $j = 1, \dots, J$, can be treated as having an approximate standard normal distribution. Graphical and numerical outlier detection procedures are based on these standardized values. We carry out the same types of analyses with these values as with the standardized ratios calculated under case 1. For formal inferences we account approximately for the correlation among terms within groups (approximately $-1/(J - 1)$), as we did in case 1, by treating the J values within each group as if they were $J - 1$ independent values

Case 3. Expected response probability is moderate (i.e. $0.1 \leq \hat{p} \leq 0.9$) but some expected frequencies do not exceed the cutoff (e.g. 5)

We follow the suggestion of Barnett and Lewis and carry out the arcsine variance stabilizing transformation. In particular $2N_j^{1/2} [\arcsin(\hat{p}_j^{1/2}) - \arcsin(\hat{p}^{1/2})]$ has an approximate standard normal distribution as $N_j \rightarrow \infty$. We estimate p by \hat{p} . It can be shown that the variance of $2N_j^{1/2} [\arcsin(\hat{p}_j^{1/2}) - \arcsin(\hat{p}^{1/2})]$ is approximately $(1 - N_j/N)$. If all the N_j 's are equal this variance is $1 - 1/J$. Thus the quantities $2N_j^{1/2} (1 - N_j/N)^{-1/2} [\arcsin(\hat{p}_j^{1/2}) - \arcsin(\hat{p}^{1/2})]$, $j = 1, \dots, J$ can be treated as having an approximate standard normal distribution. Graphical and numerical outlier detection procedures are based on these standardized values. We carry out the same types of analyses with these values as with the standardized ratios calculated under case 1. For formal inferences we account approximately for the correlation among terms within groups (approximately $-1/(J - 1)$), as we did in the previous cases, by treating the J values within each group as if they were $J - 1$ independent values.

APPENDIX AXV.* CONFIDENCE INTERVAL ON CONCENTRATION THAT CORRESPONDS TO A GIVEN LEVEL OF INCREASE IN RESPONSE OVER CONTROL GROUP RESPONSE.

After we have fitted the nonlinear regression model we wish to calculate confidence bounds on the safe dose. We use Fieller's theorem.



Suppose we're willing to tolerate a response rate L above the control group rate, C .

We want a confidence interval on d_{safe} such that

$$\Phi(\beta_0 + \beta_1 d_{\text{safe}}) = L \quad \text{and} \quad \theta \equiv (\beta_0, \beta_1, c).$$

The standard probit fit assumes that

$$p(\theta; d) = c + (1 - c)\Phi(\beta_0 + \beta_1 d). \quad (1)$$

where c is the background rate.

We obtain $\hat{\theta}$ by a maximum likelihood fit of the model using a nonlinear regression program or using SAS PROC PROBIT.

We then wish to solve the equation

$$\Phi(\beta_0 + \beta_1 d_{\text{safe}}) = L. \quad (2)$$

$$\text{Thus,} \quad \beta_0 + \beta_1 d_{\text{safe}} = \Phi^{-1}(L) \equiv f_L \quad (3)$$

*Appendix AXV is the appendix for Section XV.

The point estimate for d_{safe} is $\hat{d}_{\text{safe}} = \frac{\phi^{-1}(L) - \hat{\beta}_0}{\hat{\beta}_1}$ (4)

Placing a confidence interval on d_{safe} is now a direct application of Fieller's Theorem. See Mandel [44], page 279 or Graybill [45], pages 126-127

Thus for fixed d , a $1 - \alpha$ confidence interval on $y_d \equiv \beta_0 + \beta_1 d$ is

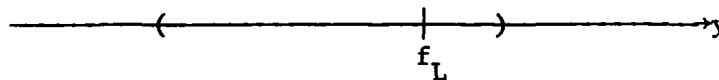
$$y_d \in \hat{\beta}_0 + \hat{\beta}_1 d \pm z_{\alpha/2} \sqrt{g + 2hd + jd^2} \quad (5)$$

where $g \equiv \text{Var}(\hat{\beta}_0)$, $j \equiv \text{Var}(\hat{\beta}_1)$, $h \equiv \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.

The confidence interval on d_{safe} includes all d 's such that

$$f_L \in \hat{\beta}_0 + \hat{\beta}_1 d \pm z_{\alpha/2} \sqrt{g + 2hd + jd^2}$$

$$\text{i.e. } (\hat{\beta}_0 + \hat{\beta}_1 d - f_L)^2 \leq z_{\alpha/2}^2 (g + 2hd + jd^2)$$



Thus the limits on d_{safe} are obtained by solving the equations

$$f_L = \hat{\beta}_0 + \hat{\beta}_1 d \pm z_{\alpha/2} \sqrt{g + 2hd + jd^2} \quad (6)$$

Thus

$$(\hat{\beta}_0 + \hat{\beta}_1 d - f_L)^2 = z_{\alpha/2}^2 (g + 2hd + jd^2) \quad (7)$$

A necessary and sufficient condition for Fieller's Theorem to yield a valid confidence interval for d_{safe} is that

$$\hat{\beta}_1^2 - jz_{\alpha/2}^2 > 0 \quad (8)$$

That is, $\hat{\beta}_1^2$ must be "many" std error units away from 0. Under the condition in equation (8) a $1 - \alpha$ two sided confidence interval on d_{safe} is

$$d_{\text{safe}} \in \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \quad (9)$$

where A, B, C are

$$\begin{aligned} A &= \hat{\beta}_1^2 - jz_{\alpha/2}^2 \\ B &= 2[\hat{\beta}_1(\hat{\beta}_0 - f_L) - hz_{\alpha/2}^2] \\ C &= [(\hat{\beta}_0 - f_L)^2 - gz_{\alpha/2}^2] \end{aligned} \quad (10)$$

Suppose now we wish to calculate a $1 - \alpha$ level lower bound on d_{safe} . This is the value that could be used for regulatory purposes. By an argument similar to the one above, it can be shown that the form of the confidence interval is the same as above, except that $z_{\alpha/2}$ is replaced by z_α and the smaller root in equation (9) is used. Thus to calculate a lower 95 percent confidence bound on d_{safe} we use the lower end point of a 90 percent two sided confidence interval and d_{safe} , etc.

APPENDIX AXVI.1.* CONFIDENCE BOUNDS ON BINOMIAL PROBABILITIES

We observe $\hat{p} = \frac{Y}{n}$ $Y \sim \text{Bi}(n, p)$.

We wish to construct a $1 - \alpha/2$ sided confidence interval on p .

These limits can be obtained easily using the Clopper-Pearson charts.

See Dixon and Massey [13] pp 501-504.

Tables of such confidence limits are given in Natrella [48].

$P \{ \underline{p} \leq p \leq \tilde{p} \} = 1 - \alpha$ where

$$\underline{p} = \left[1 + \frac{n - Y + 1}{Y} F(2n - 2Y + 2, 2Y; 1 - \alpha/2) \right]^{-1} \quad \underline{p} = 0 \text{ if } Y = 0$$

$$\tilde{p} = \left[1 + \frac{n - Y}{Y + 1} \frac{1}{F(2Y + 2, 2n - 2Y; 1 - \alpha/2)} \right]^{-1} \quad \tilde{p} = 1 \text{ if } Y = n$$

*Appendix AXVI.1 is an appendix for Section XVI.

APPENDIX AXVI.2.* DESCRIPTION AND INSTRUCTIONS FOR USE OF NONPARAMETRIC
DOSE RESPONSE ESTIMATION COMPUTER PROGRAM

*Appendix AXVI.2 is an appendix for Section XVI.

A COMPUTER PROGRAM TO CALCULATE
NONPARAMETRIC LOWER CONFIDENCE BOUNDS ON
SAFE CONCENTRATIONS IN QUANTAL RESPONSE
TOXICITY TESTS*

by

Paul I. Feder and Mary E. Sherrill
The Ohio State University

Technical Report No. 215

September 22, 1980

* This work was carried out with the support of the U. S. Army Medical Bioengineering Research and Development Laboratory, Frederick, Maryland under Contract DAMD17-79-C-9150 at The Ohio State University.

I. INTRODUCTION

Dose response experimentation has many applications such as toxicity tests, bioassay, engineering stress tests, tests of response to advertising campaigns, just to name a few. This writeup considers application of dose response experimentation to tests of toxicity on fish and other water species. However the procedures and the computer program discussed are relevant to many other applications.

In aquatic toxicity tests samples of fish or other water species are exposed in tanks or beakers to the substance or substances under study. The exposures are carried out at a succession of concentrations starting at or near zero (i.e. the control groups) and progressing to relatively high and lethal concentrations. Responses of these creatures to the toxicant are recorded and degrees of response are compared across concentration groups.

Many different responses are recorded. Some of these are percent of embryos hatched live or hatched normal, percent of fry surviving for a fixed duration or to the conclusion of the test, body weight or body length of surviving fish, numbers of eggs laid, numbers of eggs hatched. These various responses give rise to several different types of data-quantitative (e.g. body weights), count (e.g. number of eggs laid), quantal response (e.g. hatch/no hatch, live/die, normal/abnormal). The discussion here pertains to quantal responses.

Quantal response toxicity tests often give rise to binomial distribution data. Namely within each test chamber a certain number of organisms are placed on test. Under reasonable assumptions the numbers of "successes" (e.g. numbers of fish per tank that die before the conclusion of the test) follow the binomial distribution. Many standard methods have been proposed over the years to fit models to such binomial dose response data. Finney (1971) and Cox (1970) discuss two of the most commonly used empirical models, namely the probit and logit dose response functions. Background response is commonly accounted for in these models by means of Abbot's correction (see Finney, Chapter 7). A number of other parametric dose response models have been proposed, based on empirical or on mechanistic considerations. See Kalbfleisch and Prentice (1980), pp. 195-198 for a description of a general family of dose response models.

Determining safe concentrations by inferences based on dose response curves has an important advantage over determinations based on hypothesis tests to compare treatment and control groups. Namely if a particular test is either too small or too variable then a hypothesis test comparing treatment and control group response rates may not be sufficiently powerful. It may thus not be able to detect moderate sized changes in response rate from the control group. This has the effect of raising the estimated "no effect" level, which is unconservative. Such a problem might well arise in the presence of a reasonable sized background response rate. By contrast, decreased sample size or increased variability reduce lower confidence bounds on safe

concentrations derived from dose response curve fits. In this sense, inferences about safe concentrations based on dose response curves are more conservative than inferences based on hypothesis tests.

However parametric dose response models have the common drawbacks that

1. Inferences about percentiles, especially low percentiles, of the response distribution can be very sensitive to the specific functional form assumed.
2. Inference procedures are generally based on asymptotic normal maximum likelihood theory and may thus be inappropriate for data sets with small sample sizes or with many treatment group response rates near 0% or near 100%.
3. Results of the high concentration treatment groups, far away from the safe concentrations, have important influence on the estimation of the low percentiles of the dose response curve.
4. Background responses are accounted for in a structured parametric manner, such as by Abbott's correction. The form of this background correction may influence the determination of safe concentrations.

The procedure discussed in this writup avoids many of these problems. It does not require assumption of a specific functional form of the dose response curve. It is based on exact small sample distribution theory. It uses information only from the lower half of the dose response curve. It does not require a structured, parametric form of correction for background.

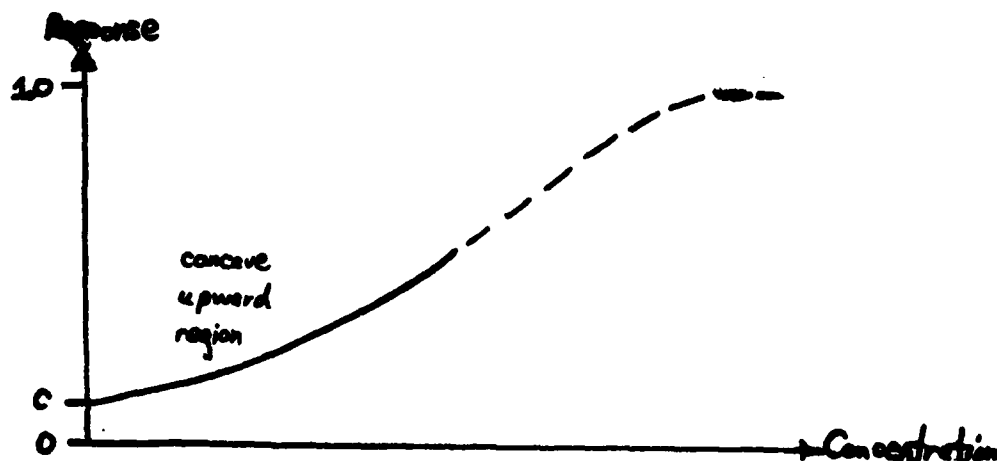
One assumption made throughout this writeup is that no tank to tank heterogeneity is present within treatment groups. This implies that the responses can be pooled across tanks within treatment groups and summarized by a single binomial distribution per group. If tank to tank heterogeneity exists it can be accounted for by adjusting the data to reflect within tank correlation, by fitting more complex models which explicitly account for such heterogeneity, or by carrying out analyses on a per tank basis. The procedure discussed in this writeup can be used in conjunction with the first adjustment approach.

II. PROCEDURE

The aim is to calculate a lower confidence bound on the "safe" concentration. The "safe" concentration is defined to be the greatest concentration for which the response rate is at most 100L% above the control group response rate. Note that this does not imply that 100L% response to toxicant is considered to be "acceptable". We wish to eliminate risk altogether. However, we can be confident with such a criterion that at worst we have limited the risk to this level.

The procedure described in this section is a nonparametric approach to determining such a safe concentration. Thus it is not necessary to specify a particular parametric form for the dose-response curve. The procedure was motivated by one discussed in Gross, Fitzhugh, and Mantel (1970) for quantitative response, but differs from it in a number of respects.

Consider a dose response curve relating percent response (e.g. mortality) to toxicant concentration.



We concentrate on the portion of the curve that is concave upward.

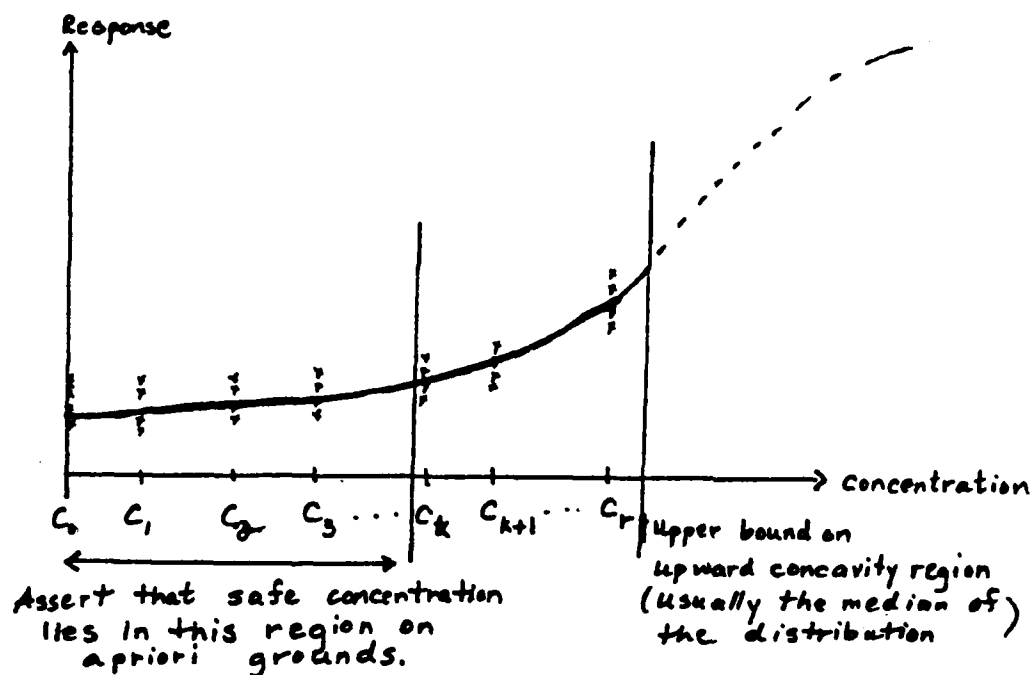
In the case of a probit or a logit model this would be all concentrations below the EC50, if there is no background and even beyond the EC50 if there is background response.

Suppose that the toxicity test involves several tanks per concentration group and that we have determined which concentration are in the concave upward portion of the curve (e.g. below the EC50). We might do this by looking at a graph or by a preliminary analysis. Let $c_0, c_1, c_2, c_3, \dots, c_r$ denote these concentrations. Since we assume the absence of tank to tank heterogeneity within groups, we pool the responses across tanks within groups to obtain the estimated response rates \hat{p}_0 at c_0 (control), \hat{p}_1 at c_1 , \hat{p}_2 at c_2 , \hat{p}_3 at c_3 , ..., \hat{p}_r at c_r . Let $p_0, p_1, p_2, p_3, \dots, p_r$ denote the "true" response rates at these concentrations.

We wish to construct a lower confidence bound on the "safe" concentration.

Definition. A safe concentration is one that increases the response at most L (limit) units above background.

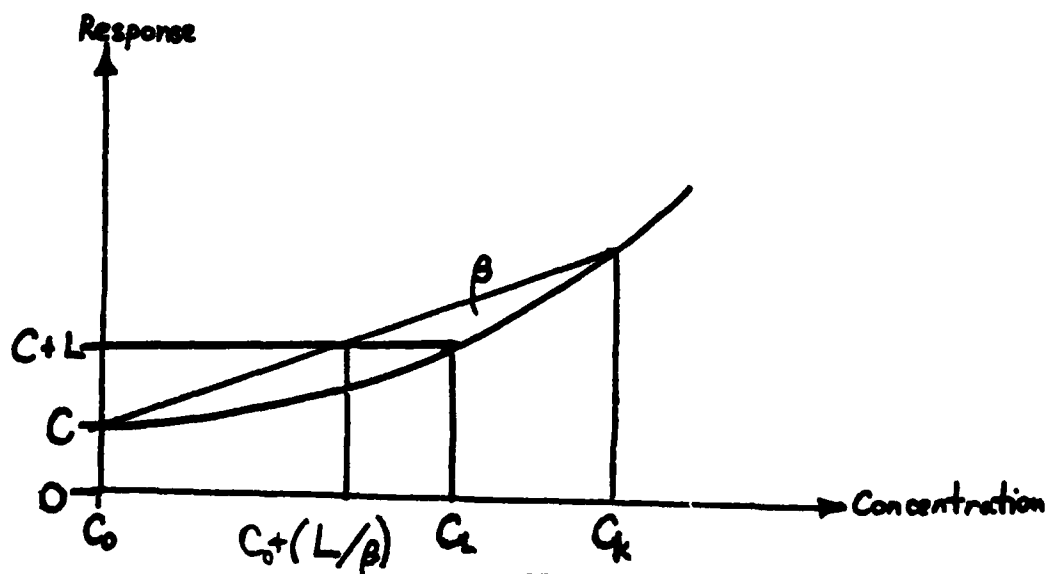
Let C_L denote this "safe" concentration. Suppose it can be asserted on a priori grounds that C_L lies in the interval $(0, C_k)$ where $C_1 \leq C_k \leq C_r$.



We construct upper bounds on the dose response curve. These upper bounds will be used to construct lower bounds on C_L . First construct a chord joining (C_0, p_0) and (C_k, p_k) . Let β denote its slope.

$$\beta = \frac{p_k - p_0}{C_k - C_0} \quad (1)$$

Concavity implies that the chord lies above the dose response curve



throughout the interval (C_0, C_k) . Thus L/β is a lower bound on C_L .

Let \tilde{p}_k be an exact upper confidence bound on p_k and let p_0 be an exact lower confidence bound on p . Expressions for such exact confidence bounds on p_0, p_k were derived by Clopper and Pearson (1934) and are contained in Hollander and Wolfe (1973), pp. 23-24. Set

$$\hat{\beta}_u = \frac{\tilde{p}_k - p_0}{C_k - C_0} \quad (2)$$

Then $\hat{\beta}_u$ is an upper confidence bound on β . Thus $C_0 + L/\hat{\beta}_u$ is a lower confidence bound on L/β and therefore also on C_L . Since we assume that $C_L \leq C_k$, our final lower confidence bound on C_L is

$$\hat{C}_L = \min(C_0 + L/\hat{\beta}_u, C_k) \quad (3)$$

It may be possible to increase \hat{C}_L by including information from $C_{k+1}, C_{k+2}, \dots, C_r$ in addition to that from C_0, C_k . Namely, we fit straight lines using $(C_0, C_k), (C_0, C_{k+1}), (C_0, C_{k+2}), \dots, (C_0, C_r), (C_0, C_k, C_{k+1}), (C_0, C_k, C_{k+2}), \dots, (C_0, C_k, C_r), (C_0, C_{k+1}, C_{k+2}), (C_0, C_{k+1}, C_{k+3}), \dots, (C_0, C_{k+1}, C_r), \dots, (C_0, C_k, C_{k+1}, C_{k+2}, \dots, C_r)$. That is, we include C_0 and all possible combinations of $(C_k, C_{k+1}, \dots, C_r)$. Thus there will be $2^{(r-k+1)} - 1$ lines calculated altogether. The lines are fitted by ordinary least squares.

Consider the calculation of an upper bound on the slope of the line based on the subset of concentrations: $C_0, C_{k_1}, C_{k_2}, \dots, C_{k_j}$ where $C_k \leq C_{k_1} \leq C_{k_2} \leq C_{k_j} \leq C_r$. Denote this as the "s-th subset" (in some ordering of subsets). Define

$$\bar{c}_s = (c_0 + \sum_{j=1}^J c_{kj}) / (J + 1) \quad (4)$$

$$\hat{\beta}_{u,s} = \frac{(c_0 - \bar{c}_s)p_0^* + \sum_{j=1}^J (c_{kj} - \bar{c}_s)p_{kj}^*}{(c_0 - \bar{c}_s)^2 + \sum_{j=1}^J (c_{kj} - \bar{c}_s)^2} \quad (5)$$

$$p_0^* = p_0$$

$$p_j^* = \begin{cases} p_{kj} & \text{if } c_{kj} < \bar{c}_s \\ \tilde{p}_{kj} & \text{if } c_{kj} > \bar{c}_s \end{cases} \quad (6)$$

$s = 1, 2, \dots, 2^{(r-k+1)} - 1$. We let

$$\hat{\beta}_u = \min_s \hat{\beta}_{u,s} \quad (7)$$

$$\hat{c}_L = \min(L/\hat{\beta}_u, c_k) \quad (8)$$

There is a problem associated with simultaneous inference. We wish to assert with a given level of confidence that $\hat{\beta}_u \geq \beta$. It can be proved that if $p_j \leq p_j \leq \tilde{p}_j$, $j = 0, k, k+1, k+2, \dots, r$ then $\hat{\beta}_u \geq \beta$. To guarantee this, with specified probability $1 - \alpha$, we use Bonferroni's inequality. Namely we calculate $r-k+2$ simultaneous two-sided $1 - \alpha$ confidence intervals by calculating each two-sided confidence interval at level $1 - \alpha/(r-k+2)$. Thus, with probability $1 - \alpha$, at least, all the intervals are assured of containing their "true" response rates.

In the case of "large" samples we can calculate alternative upper bounds on β by use of the normal approximation to the binomial distribution. Assume for definiteness that the minimum cell frequency is at least 5. That is, $N_j \hat{p}_j \geq 5$, $N_j \hat{q}_j \geq 5$, $j = 0, k, k+1, \dots, r$. (Actually Dixon and Massey (1969), page 238, state a more liberal standard.) Then

$$\hat{\beta}_s = \frac{(c_0 - \bar{c}_s) \hat{p}_0 + \sum_{j=1}^J (c_{kj} - \bar{c}_s) \hat{p}_{kj}}{(c_0 - \bar{c}_s)^2 + \sum_{j=1}^J (c_{kj} - \bar{c}_s)^2} \quad (9)$$

is a point estimate of

$$\beta_{0,s} = \frac{(c_0 - \bar{c}_s) p_0 + \sum_{j=1}^J (c_{kj} - \bar{c}_s) p_{kj}}{(c_0 - \bar{c}_s)^2 + \sum_{j=1}^J (c_{kj} - \bar{c}_s)^2} \geq \beta \quad (10)$$

and

$$\text{s.e.}(\hat{\beta}_s) \equiv \text{std.err.}(\hat{\beta}_s) = \left[\frac{(c_0 - \bar{c}_s)^2 \frac{\hat{p}_0 \hat{q}_0}{N_0} + \sum_{j=1}^J (c_{kj} - \bar{c}_s)^2 \frac{\hat{p}_{kj} \hat{q}_{kj}}{N_j}}{[(c_0 - \bar{c}_s)^2 + \sum_{j=1}^J (c_{kj} - \bar{c}_s)^2]^2} \right]^{1/2} \quad (11)$$

Thus

$$\hat{\beta}_{u,s} \equiv \hat{\beta}_s + z_{1-\alpha} \text{s.e.}(\hat{\beta}_s) \quad (12)$$

is an upper $1 - \alpha$ level normal theory confidence bound on β .

To obtain simultaneous upper confidence intervals we use Bonferroni's inequality for family size $2^{(r-k+1)} - 1$; i.e., individual intervals at level $1 - \alpha/(2^{r-k+1} - 1)$. As before

$$\hat{\beta}_u = \min_s \hat{\beta}_{u,s} \quad (13)$$

$$\hat{C}_L = \min(L/\hat{\beta}, C_k) \quad (14)$$

III. INSTRUCTIONS FOR PROGRAM INPUT

Mortality or Abnormality Data:

In this program, Subroutine RDSURV handles mortality data input. The mortality or abnormality data is read from Fortran file 4. Each input card of the mortality data must supply the program with the following two numbers (in this order):

1. The number of organisms tested in each replicate of each treatment, and
2. the number of organisms in each replicate surviving the test (or surviving normally).

The data must be inputted according to increasing order of treatments.

Important. The user supplies the input format for the card image of the survival data (see input card 4 in Fortran file 5) with variable format statements.

Concentration (Dose) Data:

In this program, Subroutine RDCONC handles concentration data input. The concentration data is read from Fortran file 9. Each input card from this file must supply the program with the following two numbers (in this order):

1. The treatment number, and
2. the concentration measurement corresponding to this treatment number.

Notes to the user:

1. The data must be inputted in increasing order of treatments.
2. The user may input many concentration measurements for a given treatment since the program calculates an average concentration for each treatment.

Important. The input format for the card image of the concentration data is user-supplied with variable format statements.
(See input card 6 in Fortran file 5).

USER'S JOB CARD

```
//
REGION=256K,MSJLEVEL=(1,1)
//JOBPARM      LINES=2000,DISKIO=2500
//SI           EXEC FORTALG,PARM.FJRT='OPT=2'
//FORT.SYSIN   DD *
```

PROGRAM DECK IS PLACED HERE

```
//
//GO.SYSLIB    DD
//              DD DISP=SHR,DSN=SYS1.IMSL.OOUBLE
//              DD DISP=SHR,DSN=SYS1.IMSL.SINGLE
//GO.FTUSF001 DD *
```

BENDIT1 - ANALYSIS OF FRY DATA

```
0 *
5
(116,15X,2F5.0)
5
112,T30,F12.4)
15.0      0.05      1
5 3 4 5
5 0.01 0.05 0.10
//
```

Information from FORTRAN
Coding Form in Figure 1.

```
//GO.FTUSF001 DD *
```

INVESTIGATOR: DUANE BENJIT (01), TEST: A (01)
DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS
SIX LEVELS: 1(CONTROL), 2(LOWEST), ..., 6(HIGHEST); 4 REPS EA (A,B,C,D)
EMBRYOS TESTED, # ALIVE AFTER HATCH, # NORMAL FRY AFTER HATCH
FRY TESTED, # ALIVE AT END, # NORMAL AT END

Five descriptive
lines at beginning
of mortality
data set

1	A	1	10101	30	24	23	15	15	15
1	B	1	10101	30	26	25	15	15	15
1	C	1	10101	30	20	20	15	15	15
1	D	1	10101	30	24	23	15	15	15
2	A	1	10101	30	22	21	15	15	15
2	B	1	10101	30	21	21	15	14	14
2	C	1	10101	30	21	21	15	15	15
2	D	1	10101	30	16	16	15	15	15
3	A	1	10101	30	17	16	15	14	14
3	B	1	10101	30	20	20	15	14	14
3	C	1	10101	30	22	21	15	15	15
3	D	1	10101	30	22	21	15	15	15
4	A	1	10101	30	12	12	15	12	12
4	B	1	10101	30	13	13	15	11	11
4	C	1	10101	30	22	22	15	14	14
4	D	1	10101	30	25	24	15	14	14
5	A	1	10101	30	20	20	15	09	09
5	B	1	10101	30	17	17	15	05	06
5	C	1	10101	30	17	16	15	09	09
5	D	1	10101	30	24	24	15	03	03
6	A	1	10101	30	19	18	15	03	03
6	B	1	10101	30	20	19	15	03	03
6	C	1	10101	30	25	25	15	12	12
6	D	1	10101	30	21	21	15	10	10

Mortality data set.
The two columns with
* at the top are
those being used in this
example, as specified in
Card #4 in Figure 1.

//GO.FT09F001 00 *

INVESTIGATOR: DUANE BENOIT (01), TEST: A (01)

DATA FROM EARLY LIFE STAGE TESTS WITH FATHEAD MINNOWS

SIX LEVELS: 1(CONTROL), 2(LOWEST), ..., 6(HIGHEST); 4 REPS EA (A,B,C,D)

SEE DATA SHEET FOR NOTES

MEASURED CONCENTRATIONS OF TOXICANT (MICRO G/L)

Five descriptive
cards at beginning
of concentration
data set

* 1 A 3 10101 061179 0.05
1 A 3 20101 062579 0.09

1 A 3 30101	070979	0.15
1 B 3 10101	060679	0.00
1 B 3 20101	061879	0.03
1 B 3 30101	070279	0.10
1 C 3 10101	060579	0.07
1 C 3 20101	062179	0.06
1 C 3 30101	070579	0.13
1 D 3 10101	061479	0.05
1 D 3 20101	062879	0.10
2 A 3 10101	061179	1.63
2 A 3 20101	062579	1.59
2 A 3 30101	070979	2.00
2 B 3 10101	060679	1.53
2 B 3 20101	061479	1.69
2 B 3 30101	061879	1.27
2 C 3 10101	070279	1.59
2 C 3 20101	060579	1.70
2 C 3 30101	061879	1.40
2 D 3 10101	062179	1.52
2 D 3 20101	070579	2.15
2 D 3 30101	061479	1.84
2 D 3 40101	062379	1.84
3 A 3 10101	061179	3.03
3 A 3 20101	062579	2.69
3 A 3 30101	070979	4.00
3 B 3 10101	060679	3.26
3 B 3 20101	061479	3.58
3 B 3 30101	061879	2.81
3 B 3 40101	070279	3.32
3 C 3 10101	062179	2.05
3 C 3 20101	061379	2.36
3 C 3 30101	062179	2.72
3 C 3 40101	070579	3.61
3 D 3 10101	061479	3.79
3 D 3 20101	062879	3.44
3 D 3 30101	060679	3.99
3 D 3 40101	061179	0.37
4 A 3 10101	062579	5.57
4 A 3 20101	070979	7.30
4 B 3 10101	060679	7.06
4 B 3 20101	060579	0.83
4 B 3 30101	062179	7.19
4 B 3 40101	061379	5.91
4 C 3 10101	062579	0.09
4 C 3 20101	070279	6.49
4 C 3 30101	070579	7.92
4 C 3 40101	070979	7.80
4 D 3 10101	060679	5.66

Concentration data set.
The two columns with
*'s at the top are those
being used in this example,
as specified in card #6 in
Figure 1.

4	C	3	20101	061179	5.21
4	C	3	30101	062179	5.23
4	C	3	40101	062179	5.14
4	C	3	50101	062579	5.44
4	C	3	60101	070279	5.95
4	C	3	70101	070579	7.32
4	C	3	80101	070979	7.50
4	U	3	10101	061179	6.72
4	U	3	20101	061479	7.27
4	U	3	30101	062579	5.93
4	U	3	40101	062879	7.76
4	U	3	50101	070979	5.42
5	A	3	10101	061179	13.40
5	A	3	20101	062579	12.00
<hr/>					
5	A	3	30101	070979	15.00
5	B	3	10101	060679	17.20
5	B	3	20101	061879	10.80
5	B	3	30101	070279	11.80
5	C	3	10101	060379	10.60
5	C	3	20101	062179	16.70
5	C	3	30101	070579	13.10
5	U	3	10101	061479	13.60
5	U	3	20101	062879	15.30
6	A	3	10101	061179	24.10
6	A	3	20101	062579	23.30
6	A	3	30101	070979	30.00
6	B	3	10101	060579	29.80
6	B	3	20101	061879	28.30
6	B	3	30101	070279	32.70
6	C	3	10101	060879	24.20
6	C	3	20101	062179	24.20
6	C	3	30101	070579	33.40
6	U	3	10101	061479	21.20
6	U	3	20101	062379	23.30
<hr/>					
/ #					
//					
<hr/>					
<hr/>					
<hr/>					

Figure 2. Sample Card Input Deck on the AMDAHL 470 V6 Computer.

DATA INPUT

Values Used by the Program:

The input cards are read from Fortran file 5. There must be 9 input cards in this file. An example input stream is illustrated in Figure 1 and a sample deck suitable for card input on the AMDAHL 470 V6 computer is illustrated in Figure 2. The following descriptions refer to the numbered cards in Figure 1.

Card #1 is the title card. The user is to choose a title or message and place it anywhere in the first 72 columns of this card. If the user does not wish to supply a title, this first card must be blank.

A typical title is illustrated on the Fortran Coding Form, line 1.

Card #2 contains the number of treatments and the number of replicate tanks within each treatment. Suppose $N1$ = number of treatments and $N2$ = number of replicates. $N1$ and $N2$ must be integer values.

If $1 \leq N1 \leq 9$, place $N1$ in column 2 of this card.

If $N1 = 10$, place $N1$ in columns 1 and 2 of the card.

If $1 \leq N2 \leq 9$, place $N2$ in column 4.

If $N2$ = a two-digit number, place $N2$ in columns 3 and 4.

(See program limitation 1.)

An example of this card is found on Fortran Coding Form, line 2. The numbers '6' and '4' indicate that there are 6 treatments with 4 replicates in each treatment.

Card # 3 supplies the program with the number of descriptive cards (lines) at the beginning of the mortality dataset. (A descriptive card may be any card at the beginning of the dataset which is not data; e.g., a typical card might contain information about the contents of the dataset.) The number on this card must be an integer from 0 to 99. Suppose NS = # of descriptive cards.

If $0 \leq NS \leq 9$, place NS in column 2 of this card.

If $10 \leq NS \leq 99$, place NS in columns 1 and 2.

Note: If the user has no descriptive cards in the mortality dataset, place a '0' in column 2 of the card.

An example of this card is found on the Fortran Coding Form, line 3. The number '5' indicates that there are 5 descriptive cards at the head of the mortality dataset.

Card #4 contains the format for the input line image of the mortality data. This format must supply the program with the following two values (in this order): The number of objects tested within a given replicate of a given treatment, and the number of these objects surviving the test (or surviving normally). Line 4 on the Fortran Coding Form contains a typical variable format statement. This format indicates to the program to tabulate to column 16 of the card, skip 15 spaces, and read the next 2 values from the card.

Card #5 contains the number of descriptive cards at the head of the concentration dataset. This card is similar to Card # 3. Suppose NC = number of descriptive cards. If $0 \leq NC \leq 9$, place NC in column 2. If $10 \leq NC \leq 99$, place NC in columns 1 and 2. An example of this card is line 5 on the Fortran Coding Form. The number '5' indicates there are 5 descriptive cards at the head of the concentration dataset.

Card #6 contains the format for the input image of the concentration (dose) data. This card is similar to Card #4. The format of this card must supply the program with the following two numbers (in this order): an integer value for the treatment number, and a real value for the concentration measurement within that treatment.

Note line 6 on the Fortran Coding Form. This variable format instructs the program to read an integer value from the first 2 columns, tabulate to column 30, and read the real value beginning in that column.

Card # 7 contains the user's choices for 3 parameter values:

1. the upper bound on the upward concavity region (UCR),
2. the level of significance, or alpha level (ALEVEL),
3. the value of a flag indicating the user's desire for simultaneous confidence intervals. (IFLG1)

The user should place UCR in columns 1-8 with a decimal point in column 4. This allows the user to specify a number as large as 999.9999. ALEVEL is placed in columns 11-16 with a decimal point in column 12. Note that this value must be a number between 0 and 1. If the user desires simultaneous confidence intervals, the value of IFLG1 must be 1. Place a '1' in column 19, if this is the case; otherwise, leave the column blank.

Note line 7 on the Fortran Coding Form. The number '15.0' indicates that the upward concavity region lies between 0 and 15.0 (0 and 15.0 are possible concentration values.) The number '0.05' indicates the level of significance, and the '1' in column 19 tells the program that a simultaneity adjustment is desired. (Thus 0.05 is interpreted as the familywise error rate.)

Card #8 contains the number of values of k to be considered in the analysis, followed by the actual k values. (Recall that the safe concentration is assumed to lie between C_0 and C_k .) The values given on this card must be integers. Suppose NK = number of k 's to be considered. If $1 \leq NK \leq 9$, place NK in column 2 of this card. If $NK = 10$, place NK in columns 1 and 2. Leave columns 3 and 4 blank. Columns 5-24 contain the actual values of k . Each value is allotted 2 columns; e.g., columns 5 and 6 contain the first value of k , and columns 7 and 8 contain the second value, etc.

If the value for k is a treatment number from 1 to 9, place the value in the rightmost column of the field. Otherwise, the value for k would equal 10 and both columns are used for this two-digit number.

Note: NK must equal the number of actual values placed on this card.

Line 8 on the Fortran Coding Form is a typical example of Card Number 8. Column 2 contains the number '3', indicating that 3 values for k are to be considered. The numbers '3', '4', '5', are these k values,

Card #9 contains the number of L's to be considered followed by the actual values for L. (Recall that L represents the incremental response rate over background, associated with the "safe" concentration.) The values on this last card are as follows: Columns 1 and 2 contain an integer value for the number of L's. Suppose NL = number of L's. If $1 \leq NL \leq 9$, place NL in column 2. If NL = 10, place NL in columns 1 and 2. Columns 3 and 4 are to be left blank. Beginning with column 5, the remaining columns are used to specify the desired values of L. Each value of L uses 6 columns; e.g., the first value appears in columns 5-10 with a decimal point in column 6. A decimal point must be placed in the second column of a given field. Note that L must be a number from 0 to 1. The user is allowed at most 4 digits to the right of the decimal point. The example of Card Number 9 is found on the Fortran Coding Form, line 9. A '3' appears in column 2

indicating that 3 values of L are to be processed. These three values are '0.01', '0.05', and '0.10'.

PROGRAM LIMITATIONS

Array Size Limitations:

1. The number of treatments may not exceed 10 and (the number of treatments) X (the number of replicates per treatment) may not exceed 40. For example, the program could handle as many as 10 treatments with 4 replicates each, or 8 treatments with 5 replicates each.
2. The user can supply this program with no more than 10 values for k (input card 8 in Fortran file 5). Similarly, the user is allowed no more than 10 values for L (input card 9 in Fortran file 5).

Output Limitations:

1. Only one title card is allowed and the user must restrict his title to the first 72 columns of the title card. (See input card 1. in Fortran file 5.)
2. Subroutine WRITE1 prints the data summary found on the first page of output.

The number of objects tested and the number of survivals are each printed in format F6.0. Therefore, any quantity greater than 99999. will not print correctly. Similarly, the concentration for each treatment is printed in F8.4 format. Numbers greater than 999.9999 will not print correctly.

DISTRIBUTION LIST

25 copies	Commander US Army Medical Bioengineering Research and Development Laboratory ATTN: SGRD-UBG Fort Detrick, Frederick, MD 21701
4 copies	Commander US Army Medical Research and Development Command ATTN: SGRD-RMS Fort Detrick, Frederick, MD 21701
12 copies	Defense Technical Information Center (DTIC) ATTN: DTIC-DDA Cameron Station Alexandria, VA 22314
1 copy	Dean School of Medicine Uniformed Services University of the Health Sciences 4301 Jones Bridge Road Bethesda, MD 20014
1 copy	Commandant Academy of Health Sciences, US Army ATTN: AHS-CDM Fort Sam Houston, TX 78234
1 copy	Commander US Army Medical Bioengineering Research and Development Laboratory ATTN: SGRD-UBD-A/Librarian Fort Detrick, Frederick, MD 21701

END

FILMED

1983

DATE